



# 权 利 要 求 书

---

1. 在接受语音输入与执行自动语音识别(ASR)的系统中调节以适应用户接口的系统,包括接收发音的装置;

利用 ASR 处理上述发音的装置,以产生上述发音的译释并确定上述译释的置信度级;以及以上述置信度级的函数有选择地调节以适应由用户请求的上述译释的校验的装置。

2. 根据权利要求 1 中确定的系统,其中设置所述处理装置以决定上述发音的至少第一和第二译释,上述译释具有各自相联系的第一和第二置信度值,以及

其中上述置信度级确定为上述第一和第二置信度值的相对大小的函数。

3. 根据权利要求 1 中确定的系统,其中所述系统进一步包括:

为上述系统的每一用户存贮信息的装置,该信息代表了以上述系统以前使用的函数计算的成功测度,以及

从上述存贮装置中检索信息和按上述成功测度值的函数调节适应上述用户接口的装置。

4. 根据权利要求 3 中确定的系统,其中所述成功测度包括上述系统的上述每个用户的以前成功率。

5. 根据权利要求 3 中确定的系统,其中上述成功测度包括与上

述每一用户的 ASR 译释相联系的以前置信度值。

6. 根据权利要求 3 中确定的系统,其中上述系统将上述成功测度与一和用户相关的门限相比较。

7. 根据权利要求 1 中确定的系统,其中上述最后说明的装置按上述译释中误差结果的函数调节适应上述校验。

8. 自动语音识别系统包括:

产生用户发音的至少第一和第二译释以及上述译释的各自第一和第二置信度值的装置,以及

用来响应上述第一和第二置信度值的相对大力进行操作的装置,以提示上述用户在接受上述第一译释为上述发音的精确表示之前。校验上述第一译释。

9. 权利要求 8 中确定的系统,其中上述系统进一步包括:用包含上述第一译释的信息提示上述用户的装置。

10. 自动语音识别系统包括

产生用户发音的译释和上述译释的置信度值的装置,以及

用来响应上述置信度值的大小进行操作的用户接口装置,用于(a)请求上述用户在接受上述译释为上述发音的精确表示之前校验上述译释,或者(b)不经校验就接受上述译释为上述发音的精确表示。

11. 根据权利要求 10 中确定的系统,其中上述系统进一步包括,存贮信息的装置,该信息指示了上述系统在译释上述用户的发音

中以前的成功,以及

响应上述存贮信息的装置,该信息用来控制上述用户接口装置。

12. 在接受语音输入和执行自动语言识别(ASR)的系统中,适应用户接口的一种方法,包括下列步骤:

接收一发音;

利用 ASR 处理上述发音,以产生上述发音的译释,并确定上述译释的置信度级;以及

按上述置信度级的函数。有选择地适应由用户请求的上述译释的校验。

13. 根据权利要求 12 中确定的方法。其中上述处理步骤包括:确定上述发音的至少第一和第二译释,上述译释具有各自相联系的第一和第二置信度值,以及

确定置信度级为上述第一和第二置信度值的函数。

14. 根据权利要求 12 中确定的方法进一步包括:

为上述系统的每一用户存贮信息,该信息代表以上述系统以前使用的函数计算的成功测度,以及

检索信息并依据上述成功测度值的函数改变用户接口。

15. 根据权利要求 14 中确定的方法,其中所述成功测度包括上述方法的上述每一用户以前的成功率。

16. 根据权利要求 14 中确定的方法,其中此成功测度包括与上述每一用户的 ASR 译释相联系的以前的置信度值。

17. 根据权利要求 14 中确定的方法。其中所述方法进一步包括将上述成功测度与一和用户相关的门限比较。

18. 自动语音识别的一种方法包括下列步骤：

产生用户发音的至少第一和第二译释，以及上述译释的各自的第一和第二置信度值，以及

响应上述第一和第二置信度值的相对值，提示上述用户在接受上述第一译释为上述发音的精确表示之前校验上述第一译释。

19. 根据权利要求 18 中确定的方法。其中上述方法进一步包括：用包含上述第一译释的信息提示上述用户。

20. 执行自动语音识别系统的一种方法包括下列步骤：

产生用户发音的译释和上述译释的置信度值，以及

适应用户接口的工作，以响应上述置信度值的大小，这是通过  
(a) 请求上述用户在接受上述译释为上述发音的精确表示之前校验上述译释，(b) 不经校验，接受上述译释为上述发音的精确表示，或者  
(c) 拒绝上述译释，并请求上述用户提供一新的发音。

21. 根据权利要求 20 中确定的方法，其中上述方法进一步包括以下步骤：存贮信息，该信息表示上述系统在译释上述用户的发音中以前的成功，以及

适应上述用户接口以响应上述存贮的信息。

# 说 明 书

---

## 利用置信度测度的自动语音识别 (ASR) 处理

本发明涉及自动语音识别 (ASR), 特别是涉及在利用自动语音识别的系统中装备的用户接口处理, 其中, 计算个人语音输入的 ASR 译释的置信度测度, 并用它来有选择地改变给予该个人语音的处理。

自动语音识别 (ASR) 系统已开始在各种应用中被广泛接受。1989 年 5 月 2 日发给 Binkerd 等人的美国专利 4827500 描述了一种在呼叫目标中选择的自动语音识别技术, 其中呼叫者与一具有 ASR 功能的声音响应单元互相配合。这种系统或者请求一口头输入, 或者为用户提供一选择菜单, 然后等待口头响应, 利用 ASR 译释该响应, 并执行所请求的动作, 所有这一切都没有人的介入。

有设计使用 ASR 的系统的用户接口中的一个重要问题, 是关系到处理识别误差的可能性的问题, 这是因为每当 ASR 系统译释一发音时, 总存在关于发音和译释之间对应性的剩余不定度。这个问题对于数字串的输出尤为重要, 例如在一系统中呼叫者说出电话号码或信用卡号码。这是因为对于数字串(而且, 在某些情况下, 甚至对于数字串的一段)仅有 85%~90% 总精确率并不鲜见。为了对付可能的误差, 现在的系统对所有的处理使用某种类型的显式校验, 其

中误差率引起关注,从而避免处理不正确数字串的可能性。比如,在每一连接的数字串输入后,ASR系统可“回读”最优数字串候选者,并需要使用该系统的个人给一肯定或否定的响应。一个例子是:“如果你的信用卡号码是XXX—YYYY,请说‘是’,否则说‘不是’”。虽然这种类型的显式校验常常是必需的和有用的,但它很麻烦,浪费时间和烦人,特别是对那些ASR系统的频繁用户,或者那些对ASR系统十分信任的用户。如果前一次请求未被识别的话其它系统请求用户重新输入一语音请求。然而,当确认识别以后,则进行静态校验过程。

根据本发明相,使用自动语音识别(ASR)技术的系统中设置了用户接口,以给出一动态过程,其中,根据ASR过程结果中的置信度的水平,对用户进行不同的处理。在一个实施例中,系统这样安排,以利用ASR系统输出的置信度的水平或程度,来区分一发音的易出错误的ASR译释和可能正确的译释。置信度可被确定为由ASR系数选择的第一和第二选择的近似分数(在下面定义)之差的函数。在该实施例中,用户接口这样安排,以使得当置信度较低时,由系统采取的显式校验步骤与当置信度较高时采用的行动不同。此外,根据误译释的“结果”以及相对于特定用户(其语音正在处理)的系统的历史性能,来给出不同的处理。在本发明的另一实施例中,在ASR系统译释一发音之后,评价出译释的置信度,并可能发生三种不同的与用户的相互作用。

示例性地，当 ASR 系统用于识别数字位时，译释的置信度可以通过在每一发音数字与每一 ASR 系统已训练的数字模型之间指定一近似记分来确定，这里大分数表示好的对应。因此，对每一发音产生一向量，表明该发音与每一模型的近似。当具有最接近度的模型的近似记分比下一个最优选择的近似记分大很多时，我们说存在着高置信度。这本质上意味着该译释大大优于其它译释。

通过将 ASR 系统性能结果的置信度式“确定性水平”变换为几个不同的动作，这些动作由产生误差结果的详细分析以及与用户响应校验请求和/或重新输入信息的困难的详细分析共同确定，系统的用户接口得到大大改善，当这些动作被断定时，用户只需重新输入或校验语音输入。

通过结合附图考虑下面的详细描述，可以更充分地理解本发明。附图中，

图 1 是一流程图，示出当一人用声音输入来拨电话号码后，常规 ASR 系统采取的步骤；

图 2 和 3 合起来是一流程图，示出依据本发明安排的 ASR 系统，对于用声音输入拨动电话号码者作出响应而采取的步骤；

图 4 是一框图，示出设置声音处理单元以实现动态用户接口过程（例如图 2 和 3 中所述的过程）的一种安排。

图 5 是一流程图，示出依据本发明的 ASR 系统中采取的步骤，其中三个不同的结果来自于产生了三种可能置信度级的 ASR 处理。

首先参考图 1，流程图示出常规 ASR 系统采取的步骤。在这个例子中，一人用声音输入拨动电话号码，ASR 系统译释该人的发音并采取行动，比如完成电话呼叫，以响应从 ASR 系统获得的发音的译释。更具体地，描述了包含拨动 10 位电话号码（3 位区域号，后面 7 位是本地电话号）的处理。

图 2 的过程由步骤 101 开始，此时呼叫者连接到一语音处理平台（下面结合图 4 描述）。该平台提供可听提示，(prompts) 接收语音输入并利用 ASR 技术译释语音。在步骤 103，用户由可听通知提示，通过在步骤 105 说出 3 位数输入电话呼叫的区域号。在步骤 106，执行任意已知的自动语音识别过程，并确定由呼叫者说出的数字。一般地，由 ASR 过程完成的译释典型地包括用户输入语音与存贮的语言样本的比较。然而，ASR 系统可实现几种熟知的语音识别过程中的任一种。

在步骤 106 识别了区域号的 3 位数字之后，系统在步骤 107 请求呼叫者显式校验被识别的数字确实与用户在步骤 105 说的数字相同。用户则在步骤 108 回答“是”或“不是”，系统则依据回答在分支步骤 111 采取不同的行动。具体说，如果在步骤 108 接收的答案为“是”，这表示头三位数字被正确识别，则过程继续前进至步骤 113，用户被提示输入电话号码其余的 7 位。用户在步骤 115 说出这七位数字，在步骤 116，如同在步骤 106 一样，再次利用 ASR 过程确定呼叫者说出的数字。接着，在步骤 117，请求呼叫者显式校验被

识别的数字与在步骤 115 说出的数字一致。如果在步骤 119 说“是”，则在分支步骤 121 识别出正响应，而系统继续，在步骤 123 利用所有 10 位被识别数字，完成这一次处理。

如果在步骤 108 或 119 从呼叫者接收到一个负响应，则该响应使得分支步骤 111 或 121 分别转移控制到步骤 125 或 127，在那里决定是否已进行太多次失败的尝试。这可以这样完成，即当过程开始时初始化一计数器，每次在步骤 111 或 121 遇到“不是”回答时将计数器加 1，并将计数器中的计数值与一预定门限相比较。如果在步骤 125 或 127 指示负响应，表示门限尚未被超出，则过程可重复，即执行步骤 103—111 或 113—121 再作识别尝试。如果在步骤 125 或 127 指示正响应，则自动语音识别已“失败”，呼叫者可在步骤 126 或 128 被连接到一人工话务员。

图 1 中所示的过程对用户产生相同的处理，即不管在步骤 106 和 116 完成的语音识别的置信度是多少，也不管与同一用户以前校验尝试相联系的历史细节如何，用户与系统间的对话是相同的。这种烦琐的、静态的方法被本发明消除了，本发明支持动态方法，它利用与在步骤 106 和 116 完成的语音识别相联系的置信度级来改变给予用户的处理。

特别地，现在参考图 2 和 3，这里示出一流程图，说明了依据本发明安排的 ASR 系统中，为了响应用声音输入拨动电话号码的人而采取的步骤。在这个示范性过程中，执行了与上述相同的处理，即

包括拨 10 位电话号码 (3 位区域号, 后面 7 位本地电话号) 的处理。该过程在步骤 201 开始, 此时呼叫者被连接到执行与上述相同的功能的语音处理平台, 而且提供与正在进行的识别相联系的置信度级的指示。置信度级确定的细节将在下面更详细地描述。下面一篇文章描述了产生与自动语音识别系统相联系的置信度测度的一种示范性技术。文章题目是“识别索引: 词汇诊断的一种统计方法”, 作者 K. P. Avila 等, 发表在 1987 年 10~11 月“speech Technology”第四卷第 1 期第 62—67 页上。

在步骤 203, 用户由一可听通知提示, 通过在步骤 205 说出了 3 位数字而输入电话呼叫的区域号。在步骤 206, 执行自动语音识别过程, 并确定由呼叫者说出的数字。象以前一样, 由 ASR 过程执行的译释典型地包括将用户输入语音与存贮的语音样本进行比较。然而, ASR 系统还提供一置信度值, 它是与识别相联系的置信度级的指示。如图 2 所示, 在步骤 231 完成的置信度分析有两个输出, 标为“很高置信度”和“中等置信度”。如下面结合图 5 所说明的那样, 可以使用多于两个的置信度级, 并且各级的定义可以不同。

如果在步骤 231 确定的置信度级是“中等置信度”, 过程以与上面所描述相同的方式继续。特别地, 系统在步骤 207 请求呼叫者显式校验被识别的数字确实与在步骤 205 用户说出的数字相同。用户则在步骤 208 回答“是”或“不是”, 系统则依据该回答在分支步骤 211 采取不同的行动。特别地, 如果在步骤 208 接收的回答为“是”,

表示头三位数字被正确识别,则过程继续前进至步骤 213,用户被提示输入电话号码的其余七位。用户在步骤 215 说出这七位数字,在步骤 216,如同在步骤 206 那样再次利用 ASR 过程,确定由呼叫者说出的数字。然而,如同在步骤 231,ASR 系统提供与识别相联系的置信度级的指示。如图 3 所示。在步骤 233 完成的置信度分析可有两个输出,标为“很高置信度”或“中等置信度”。如果步骤 233 的输出代表“中等置信度”,则请求呼叫者在步骤 217 以显式校验被识别的数字与在步骤 215 说出的数字相同。如果在步骤 218 说出“是”,则在分支步骤 221 识别出正响应,系统继续在步骤 223 利用所有十位被识别数字,完成这一处理。

在与图 1 中所用方法相似的方法中,注意如果在步骤 208 或 218 从呼叫者接收到页响应,该响应会使得支枝步骤 211 或 221 分别转移控制到步骤 225 或 227,在这里确定是否已进行了太多次失败的尝试。如果在步骤 225 或 227 指示负响应,表示门限尚未被超过,则过程可被重复,即执行步骤 203—211 或 213—221 来再次进行识别尝试。如果在步骤 225 或 227 指示正响应,则自动语音识别已“失败”,呼叫者可在步骤 226 或 228 被连接到一人工话务员。

如果在步骤 231 或 233 完成的置信度分析指示识别具有“很高置信度”,则对用户给予不同的处理。特别地,如果头三位以很高的置信度识别,则略过步骤 207,208 和 211,从而对在语音识别中关于头三位作出的判决不进行显式校验,然后,如果接下来的七位也

是以很高置信度识别，则步骤 217, 218 和 221 被略过，从而对在语音识别中关于后七位作出的判决不进行显式校验。因此，可以看出图 2 和 3 所示的过程是自适应的，它在用户和系统间产生不同的对话，该对话依赖于在步骤 206 和 216 完成的语音识别的置信度级。

如图 4 所示，一典型的语音处理单元 301 可设置用于电信网络的范围内，如 1990 年 5 月 1 日颁发给 A. N. Daudelin 的美国专利 4922529 中的图 1 所示（这里收入作为参考）。语音处理单元 301 包括通信接口 311，通过中继线 315 将其连接到其它的系统部件。接口 311 和中继线 315 可以支持多路同时双向会话，从而在任一给定时间可以处理多个呼叫。语音处理单元 301 中执行的过程由中央处理单元（CPU）303 控制，而它又在存贮器（例如数据库 309）含有的存贮程序的控制下工作。语音处理单元 301 所提供的功能度包括 (a) 利用语音发生器 307 向用户播放通知的能力，以及 (b) 利用 ASR 模块 305 译释接收的用户发音的能力。本自语音发生器 307 的通知的排序和在 ASR 模块 305 中完成的识别操作一起组成了用户接口，它依据本发明动态受控。语音处理单元的元件互相连接，并通过公共总线 313 与通信接口 311 互连。

如上所述，来自 ASR 模块 305 的输出包括：接收的用户发音的译释，以及译释中置信度的指示，后一信息可被 CPU 303 得到，从而用户接口过程可以根据置信度级动态调整。

语音处理单元 301 可利用 AT&T 的 Conversant MAP 100voice

Responso Unit 实现,它与语音识别包成套装配在一起,而存贮在数据库 309 中的控制软件可利用交互工具(称为 Script Builder)来产生。然而,要注意图 4 的语音处理单元 301 的特别设置仅仅作为示例性说明之用,而其它方法,例如在 Daudelin 专利中引用的文献里描述的方法,对于熟练的技术人员是显而易见的。特别地,应该理解:虽然结合图 1,2 和 3 描述的过程涉及进行电话呼叫范围内语音识别的应用,语音识别也可用于某“局部”过程,例如当用户与计算机或其它设备交互作用的时候。洗碗机或个人计算机可以通过在设备中加入自动语音识别单元而响应口头命令。依据本发明,计算机可以设置成响应口头发出的“格式化”命令的识别,而将磁盘格式化。由于格式化是一次可能引起数据丢失的重要操作,只有当该命令的非常高的置信度识别时,该命令才被执行。如果置信度级是中等,可以请求用户说“是”或重复该命令来显式校验该命令。如果置信度级是低的,可能需要用户在键盘上打出该命令。在这种局部情况下,通信接口 311 可被连接到一语音输入装置(如麦克风)以及输出装置(如扬声器或显示板)。

现在参考图 5,说明采用不同用户接口过程的本发明的另一实施例。在该实施例中,用户在步骤 400 被提示语音输入,ASR 模块 305 在步骤 401 译释用户发音之后,在步骤 403 确定译释的置信度,并依据三个可能的级别来评价置信度,然后与用户可能发生三种不同的相互作用。首先,如果译释正确的可能性较大,则在步骤 405 达

到一正结果，ASR 译释不经显式校验，在步骤 407 被接受，而忽略有错的可能性。处理则在步骤 409 完成。第二，对于中级不确定性，在步骤 411 达到一正结果，因此用户需要在步骤 413 显式校验（或反校验）结果，因为这可能比强迫用户重新输入信息（用声音或其它）更为有利。如果结果被验证，则在步骤 415 产生一正结果，处理在步骤 409 完成。如果结果未被验证，则在步骤 415 产生一负结果，只要没有发生太多失败的尝试，如步骤 417 确定的那样，用户就需要重复该过程（从步骤 400 开始）。第三，不确定性很大，和/或误译释的后果很严重，则步骤 405 和 411 的结果均为负。该条件作为“译释失败”处理，而用户需要“再试一次”，而不进行（可能的）错误结果的显式校验。只要用户尚未失败太多次，如在步骤 417 指示的那样，就可通过重复在步骤 400 开始的过程来实现。如果失败太多，则图 5 的过程在步骤 419 结束，因此用户（在电话呼叫的范围之内）可被连接到一人工话务员。

图 1 和 3 的步骤 231 和 233 中完成的置信度分析以及图 5 的步骤 405 和 411 中完成的置信度分析，可以这样完成：为每一发音数字对于每一已经训练的数字模型指定近似记分，大的分数表示好的对应，小的分数表示差的对应。这种方法为每一说出的数字产生了一置信度值向量，它表示了发音与每一模型的近似程度。我们已观查到，具有最大近似的选择更有可能是正确选择。这是在当次最大近似的置信度值的大小比它小很多的时候。因此，利用这两个近似

分数之差的函数来确定其发音的“最佳”译释选择确实是“正确”的选择的置信度级。置信度级的确定可以用很多方法来完成，所有方法都利用来自 ASR 系统的特定数据将可能正确的发音与正确可能性较小的发音区分开来。从这个观点看。一特定的误差率可看作由这样一个域导出，它包括  $x\%$  的有  $<a\%$  误差的数目（可看作不太容易出错）和  $y\%$  的有  $>b\%$  和  $<c\%$  误差的数目（一个更易出错的集合），以及  $z\%$  所有  $>c\%$  误差的数目（一个被认为不可能正确的集合）。可以利用以 ASR 系统和已知语音样本作的实验来确定哪些特定值用作  $x$ ,  $y$ ,  $z$  和  $a$ ,  $b$ , 以及  $c$ 。

这里还应注意，语音识别工作的两个可能输出的相对“近似”可以用不同方式来表示。可以利用分数的比率或线性差，或者利用一些更复杂的函数。最佳“近似”的具体确定依赖于所使用的具体模型的性质以及计算相似测度的算法，还可能涉及到其它变量。

依据本发明，历史细节，如与同一用户以前的校验尝试相连系的成功测度，可以用来动态改变或调整 ASR 过程和 ASR 系统与用户相互作用的方式，这是因为 ASR 系统的所有用户并不都经历同样的成功级。也不都产生同样的置信度级。可以用符号“绵羊”和“山羊”来描述这一方案，即对于某些人（即“绵羊 (sheep)”）工作得较好的 ASR 过程与对于其他人（即“山羊 (goat)”）工作得较差的过程是不同的。很清楚，当 ASR 系统将显式校验措施引入用户接口时，它改进了对于山羊的系统性能，允许较少的误差发生。同时，对

于所有用户，通过引入额外的相互作用，它降低了接口的质量，并且绵羊（其语音一般为系统所理解）对该措施的需求较小。

使用历史的“成功测度”能适应两种类型用户，因为“成功测度”能够允许可能是绵羊的用户和可能是山羊的用户之间的差异。当 ASR 处理应用于基于用户的服务时，（在一段时间内涉及同样的用户），确定或预测哪一个为“ASR 绵羊”是可能的。在这种服务中，对于某一给定用户，进行或不进行显式校验都很容易跟踪 ASR 系统返回器—高置信度分数的频度和/或—特别用户成功的频度。一贯接收高置信度分数和/或一贯成功的用户，“被假定为绵羊”。对于这些用户，即使在某些情况下置信度级不是“很高”，校验措施也可省去。的确，对于那些 ASR 系统历史上工作很好的人，中等置信度级就可以使过程跳过显式校验，并省去图 2 和 3 中的步骤 207, 208 和 211 和/或步骤 217, 218 和 221，省去图 4 中的步骤 413 和 415。对于具有大成功测度的用户，只有当步骤 231 或 233 中的结果产生“低”置信度级，或当步骤 405 和 411 的结果均为负时，才执行那些步骤。注意在某些不能获得历史信息的实现中，例如当一新用户利用声音命令操作计算机时，比较历史的用户发音与 ASR 识别，以及跟踪识别成功的频度是行不通的。

区分各种类型用户所需的历史信息可以存贮在图 4 的数据库 309 中，并可以被检存以响应个人访问语音处理单元 301。例如，可以利用自动号码识别 (ANI) 信息来识别用户，当电话呼叫从起始交

换机服务的一电话站产生时，该信息被送给该起始交换机。或者，可以利用个人识别号码 (PIN) 来识别用户，该号码由用户作为 ASR 过程的一部分给出。在上述两种情况下，ANI 或 PIN 被用作一检索关键词来确定数据库中的信息，指示是否该特定用户是要改变过程的用户，以及，如果是，如何改变。本质上，系统可以这样确定用户是绵羊还是山羊。

下面的试验对本发明进行了模拟，它的 10 位电话号码有两部分组成，3 位区域号和 7 位当地号，利用 AT&T Conversant 系统上的自动语音识别。在这个实验中，利用数字串候选者的置信度测度来改进用户接口，从而当第一数字串候选者深到比第二数字串候选者高得多的置信度分数时，不进行显式校验。特别地，AT &T Coversant 系统为多达四个可能的数字串候选者的每一个指定一在 1 和 1,000,000 之间的置信度值。具有最高置信度值的候选者称为“第一候选者”；具有次高置信度值的候选者称为“第二候选者”；依此类推。系统计算第一和第二候选者的置信度值之差，从而确定 ASR 结果的置信度级，然后利用该差别，依据哪些显式较验提示是否实行，以及过程中哪些步骤省略来调整整个过程。如果 1 号候选者和 2 号候选者之间的差别大于 6000，则假定置信度很高，足以改变过程并略过显式校验步骤。在那些置信度分数差小于 6000 的处理中，则有下面类型的对话，其中 S 代表系统提高，U 代表用户输入：

S：现在请说出您想呼叫的区域号。

U: 908

S: 您是说 9, 0, 8 吗?

U: 是

S: 现在请说出您想呼叫的 7 位电话号码。

U: 9496510

S: 您是说 9, 4, 9, 6, 5, 1, 0 吗?

U: 是

S: 谢谢!

另一方面, 如果置信度分数差大于 6000, 则有下面类型的对话:

S: 现在请说出您想呼叫的区域号。

U: 908

S: 现在请说出您想呼叫的 7 位电话号码。

U: 9496510

S: 谢谢!

ASR 性能和收集的择优数据表明, 动态利用置信度分数来调整校验过程的用户接口优于常规的用户接口。完成电话号码处理的平均时间减少约 25%; 用户偏爱使用置信度记分的系统; 且“错号”呼叫的百分比没有增加。对于其它根据置信度记分的过程调整, 也会观察到相似的结果。

关于在处理 ASR 采样中确定用户接口时利用历史数据作为成功测度, 问题可分为两组。一组, 侧定的绵羊; 被定义为那些识别

器在至少 60% 的处理中（用户进行多达 32 次 ASR 尝试）具有高置信度的用户。另一组，侧定的山羊，是剩余的用户。对于每一用户组，将总 ASR 准确度与那些识别器有“高置信度”（定义为置信度差的分数  $> 6000$ ）处理的准确度相比较。发现总的 ASR 性能表明成功率为 83.8%，然而，如果仅考虑那些 ASR 置信度高的处理，成功率为 97.5%，这表明在这些情况下，不象以前注意到的那样需要用户证实结果。然而，也可以认为识别器的准确度仅是对于“高置信度处理”期间的 ASR 假定的绵羊。数据表明，对于这些用户，ASR 系统达到极高的性能，407 次尝试中 406 次成功，准确率为 99.8%。

简言之，这些实验表明对于一些用户，识别器常常具有高置信度。对于这些个人，当置信度高时，识别器实际上总是正确的。在那些假定绵羊可被识别的情况下，可以定义最佳 ASR 用户接口为能够与人工话务员讲话一样快或比其更快完成处理的用户接口。这可能需要根据识别器置信度分数和/或用户表明的 ASR 选择或系统使用历史来进行实时呼叫流程判决。总的关键在于，用户接口应当识别山羊和绵羊的不同需求。现行大多数系统都只为山羊最优化，而本系统可对于绵羊和山羊的呼叫流程都进行最优化。

那些本领域的普通技术人员可对本发明进行各种更改。相应地，本发明仅由所附权利要求书所限。

# 说明书附图

图 1

1/5

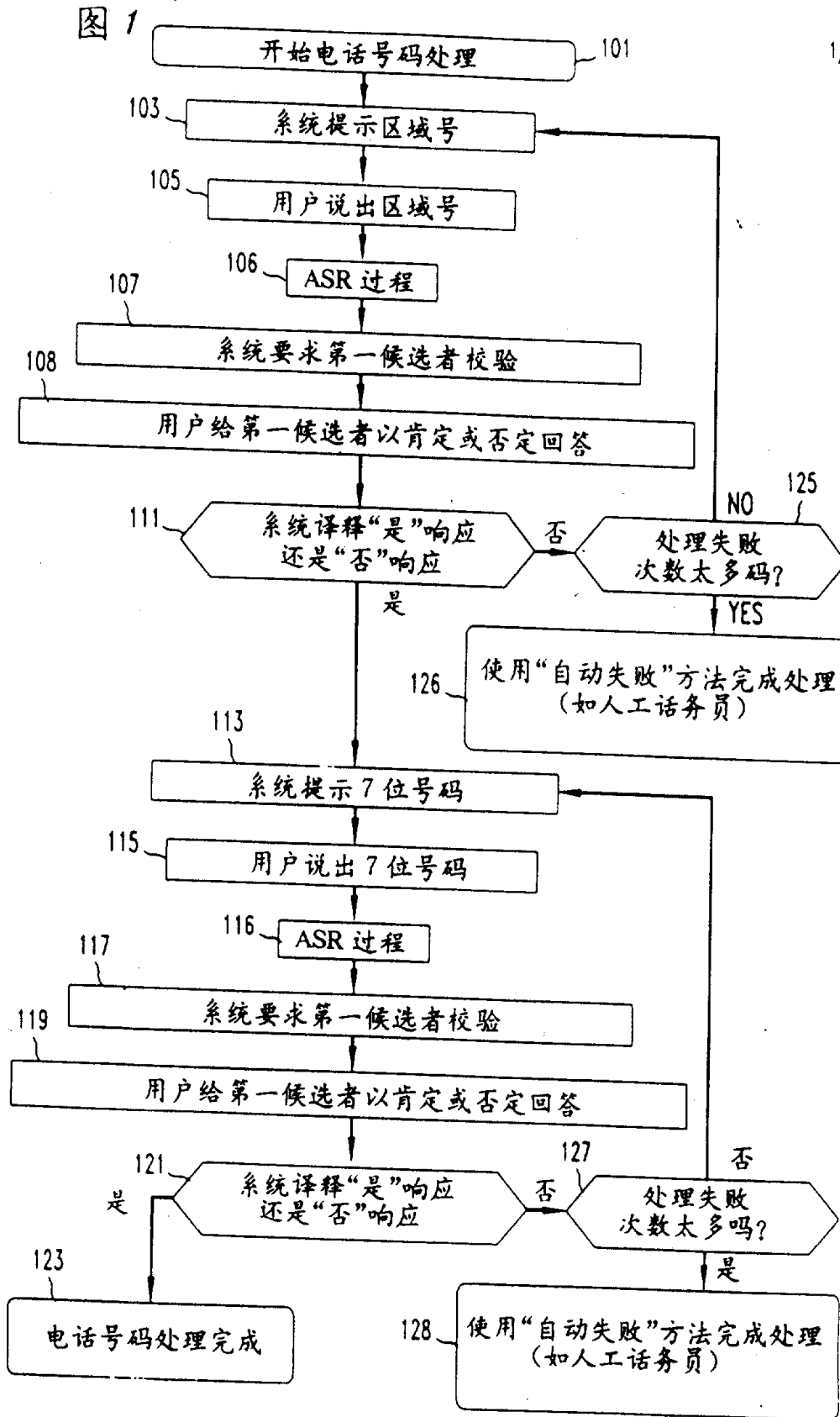
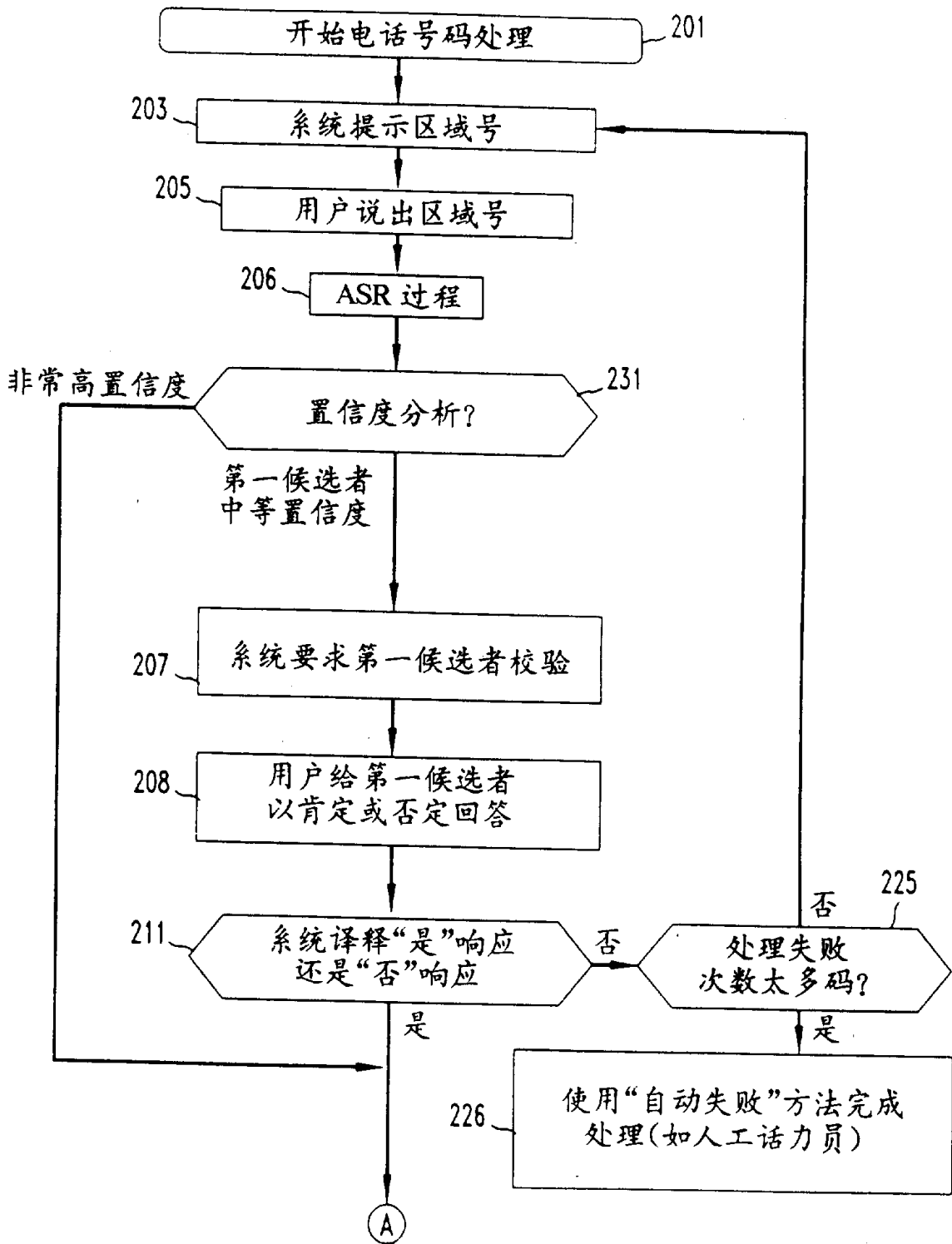


图 2



至图 3

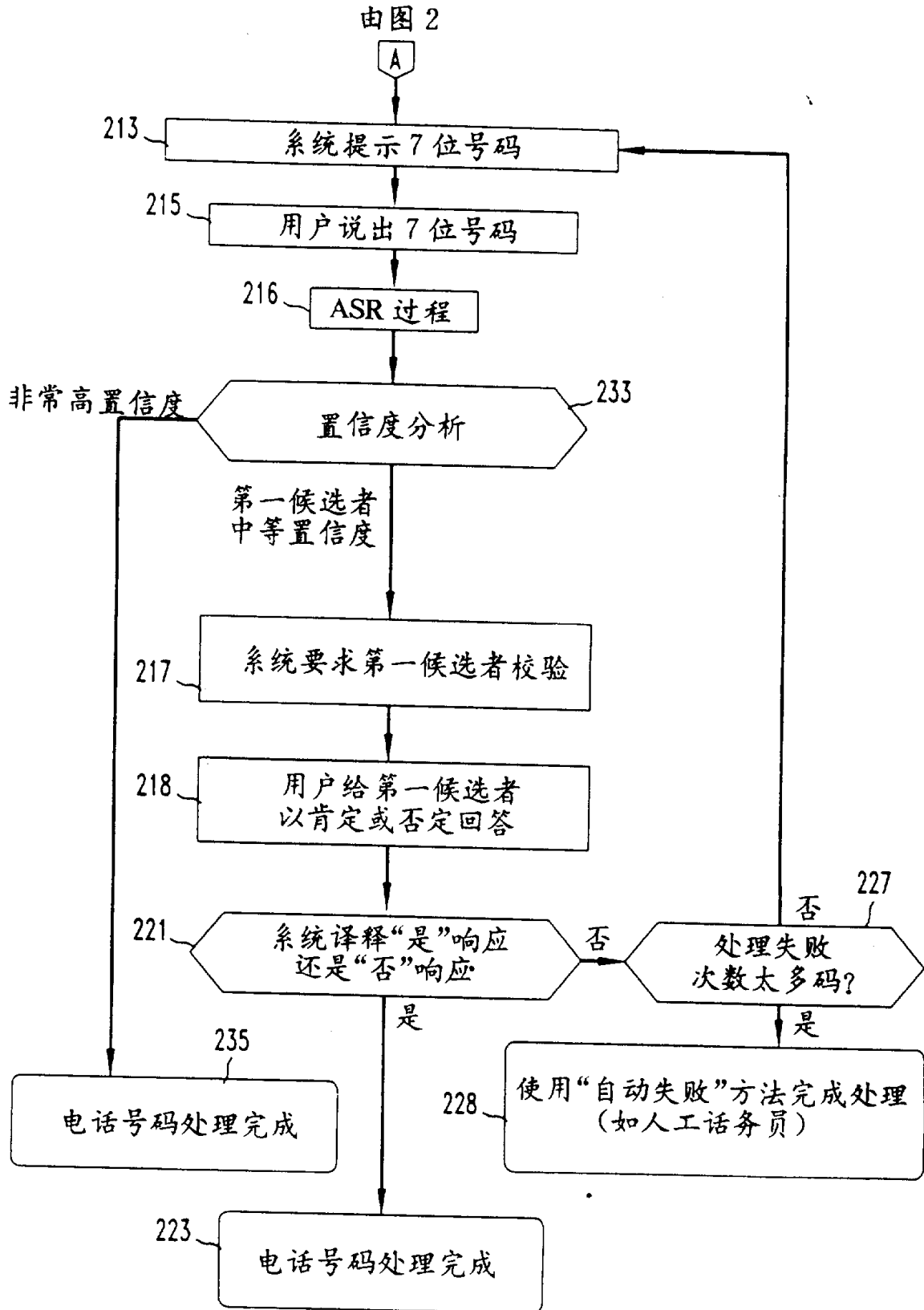


图 4

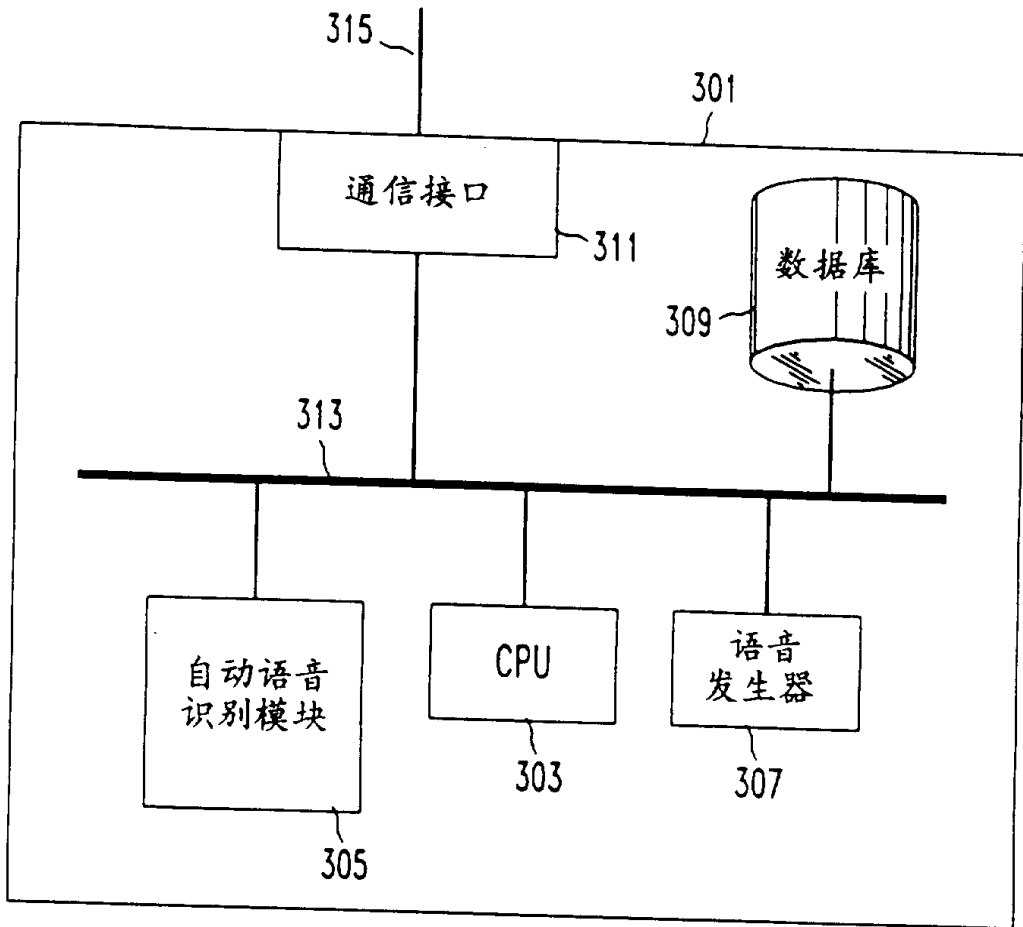


图 5

