

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 15/20 (2006.01)



[12] 发明专利说明书

专利号 ZL 200380102529.4

[45] 授权公告日 2009 年 8 月 5 日

[11] 授权公告号 CN 100524458C

[22] 申请日 2003. 10. 24

[21] 申请号 200380102529.4

[30] 优先权

[32] 2002. 11. 2 [33] DE [31] 10251113.6

[86] 国际申请 PCT/IB2003/004727 2003. 10. 24

[87] 国际公布 WO2004/042698 英 2004. 5. 21

[85] 进入国家阶段日期 2005. 4. 29

[73] 专利权人 皇家飞利浦电子股份有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 A·库伊曼

[56] 参考文献

CN1242553A 2000. 1. 26

US2002/0019734A1 2002. 2. 14

US6336091B1 2002. 1. 1

CN1300417A 2001. 6. 20

EP1085501A3 2002. 1. 9

WO/0072307A1 2000. 11. 30

Intelligent Barge - in In Conversational Systems. N. Strom and S. Seneff. Proceeding of The International Conference On Spoken Language Processing. 2000

审查员 苗 浩

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 刘 红 王忠忠

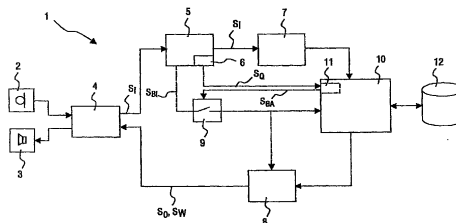
权利要求书 2 页 说明书 8 页 附图 2 页

[54] 发明名称

用于操作语音识别系统的方法

[57] 摘要

描述一种用于操作语音识别系统的方法，在该方法中检测和分析用户的语音信号(S_I)，以识别包含在该语音信号(S_I)中的语音信息。该语音识别系统确定代表当前接收质量的接收质量值(S_Q)或者噪声值。当接收质量值(S_Q)降到接收质量阈值以下时或者当噪声值超过噪声阈值时，该语音识别系统(1)转换到对噪声不太敏感和/或向用户输出报警信号(S_W)的操作模式。还描述了一种合适的语音识别系统。



1. 一种用于操作语音识别系统(1)的方法,在该方法中,检测和分析用户的语音信号(S_1),以识别包含在该语音信号(S_1)中的语音信息,其特征在于,确定代表当前接收质量的接收质量值(S_0)或噪声值,并且当接收质量值(S_0)落到给定的接收质量阈值以下时,或者当噪声值超过噪声阈值时,停用该语音识别系统(1)的插嘴操作模式。

2. 根据权利要求1所述的方法,其特征在于,当接收质量值(S_0)再次超过接收质量阈值,或者当噪声值再次落到噪声阈值以下时,该语音识别系统被自动复位到前一操作模式。

3. 根据权利要求1或2所述的方法,其特征在于,当接收质量值落到接收质量阈值以下或者噪声值超过噪声阈值时,该语音识别系统(1)输出报警信号(S_w)给用户。

4. 根据权利要求1或2所述的方法,其特征在于,利用话音活动检测器(5)来确定接收质量值(S_0)或者噪声值。

5. 根据权利要求1或2所述的方法,其特征在于,根据在用户的发声开始之前和/或在语音暂停中接收的背景信号,确定接收质量值(S_0)或者噪声值。

6. 根据权利要求1或2所述的方法,其特征在于,该话音活动检测器(5)将接收质量值(S_0)或噪声值本身施加给对话控制装置(10),和/或当接收质量值落到接收质量阈值以下时或当噪声值超过噪声阈值时,将接收恶化指示信号(S_{EB})施加给对话控制装置(10)。

7. 根据权利要求1或2所述的方法,其特征在于,当接收到接收恶化指示信号(S_{EB})和/或当接收到的接收质量值(S_0)降低到接收质量阈值以下或噪声值超过噪声阈值时,该对话控制装置(10)开始向用户输出提示(S_v),由此给予该用户接收条件差的信息。

8. 根据权利要求1或2所述的方法,其特征在于,根据引起接收质量值(S_0)低于接收质量阈值或者噪声值高于噪声阈值的干扰类型,分析输入信号,并且向用户输出包含此信息的提示(S_v)。

9. 一种语音识别系统(1),包括用于检测用户的语音信号(S_1)的装置(5)和用于分析检测的语音信号(S_1)以识别包含在该语音信号中的语音信息的语音识别装置(7),其特征在于,该语音识别系统包括:

质量控制装置(6),用于确定代表当前接收质量的接收质量值(S_0)或噪声值;

比较器,用于将接收质量值(S_0)与预定的接收质量阈值进行比较,或者用于将噪声值与给定的噪声阈值进行比较; 和

控制装置(9, 10),用于当接收质量值降到接收质量阈值以下时或者当噪声值超过噪声阈值时停用该语音识别系统(1)的插嘴操作模式。

用于操作语音识别系统的方法

技术领域

本发明涉及一种用于操作语音识别系统的方法，在该方法中，检测和分析用户的语音信号，以识别在该语音信号中包含的语音信息。本发明还涉及一种语音识别系统，该系统包括用于检测用户的语音信号的装置以及用于分析检测的语音信号以识别包含在语音信号中的语音信息的语音识别装置。

背景技术

语音识别系统被日益用于许多应用领域中，以便能够在用户与相关应用之间进行语音通信，例如设备控制或自动信息服务。这样的语音-用户接口具有许多优点。一方面，一般来说，它们使设备的免提操作成为可能。另一方面，它们还特别适合作为用户接口用于部分视力或其它身体障碍的人，这些人不能阅读光显示器和/或操作输入装置如键盘或鼠标，并因此只得依赖语音作为通信手段。此外，这种类型的接口被日益用于自动语音对话系统中。这样的语音对话系统的示例是目前特别被某些大公司和公共服务使用的自动应答和信息系统，以便向主叫者尽可能快速地和尽可能舒适地提供期望信息或者把主叫者连接到负责该主叫者的特定请求的台站。这方面的进一步实例是自动电话信息系统、导航系统、具有一般议程主题或电影和戏剧节目的信息服务以及各种信息系统的任意组合。

根据其中使用了语音识别系统的相关应用，以不同方式或经由不同设备从用户传送语音信号给语音识别系统。在固定对话系统的情况下，例如在如火车站、机场、博物馆等的公共场所的信息终端中，语音识别系统包括连接适当的麦克风、耳机或类似物的一个或多个接收信道，以使用户能够与终端或者语音识别系统通信。然而，通过使用用户的适当终端，经由通信网如电话网或者因特网，通常可以接入上述的自动语音对话系统。因此，在此情况下，用户的语音信号首先被终端检测，在此之后将其转换成电信号，并经由通信网的语音和/或数据信道将其施加到语音对话系统的终端。

特别地用户借助于通信终端从任意环境可以联络的语音识别系统的固有特性是：其中用于接收语音信号的接收质量可以强烈地改变，即，甚至在会话内强烈变化。随后，一方面利用传输信道如电话链路的质量，以及另一方面利用其本身依赖于用户在相关时刻所在的环境的背景噪声电平，将接收质量确定到一个实质范围。显然，当接收质量较低时，语音识别系统的操作条件也将不太有利，因为传输信道的噪声或者短暂干扰或中断易于破坏识别结果。当接收质量进一步恶化时，语音识别系统迟早（这取决于语音识别系统的灵敏度）将几乎不可能传送在某一程度令人满意的识别结果。对用户的发声尽可能快速地和自然地进行响应并因而对于用户来说本身是非常舒适的语音识别系统对干扰的反应非常敏感，这正是其缺点。另一方面，例如允许用户仅仅在给定的确切定义的时刻输入语音发声的语音识别系统在语音输入期间需要给定的相当高的音量，并且对发声响应相当慢，从而在接收质量恶化方面更加强烈。

发明内容

本发明的目的是以这样的方式改善上述类型的用于操作语音识别系统的方法或者语音识别系统，即，在接收质量恶化的情况中，也以尽可能高的质量继续操作，并且仍然给用户最大可能的舒适。

本发明的目的是这样实现的：确定代表当前接收质量的接收质量值或者噪声值，并且当接收质量值落到给定接收质量阈值以下时，或者当噪声值超过噪声阈值时，该语音识别系统转换到对噪声不太敏感和/或向用户输出报警信号的操作模式。

由于除了使用倒数值以及限制条件的相应颠倒之外，在接收质量值降低到给定的接收质量阈值以下方面监视接收质量值等同于在噪声值（例如背景噪声信号的电平）超出给定的噪声阈值方面监视噪声值，因此为了简便起见，在一般情况下，下面将只根据第一形式来描述本发明，然而，这不以任何方式限制本发明。在下面的说明中，这两个形式的相应术语总是可以进行互换。

本发明使语音识别系统本身和/或用户能够对接收质量恶化进行适当响应，所以用户总可以在当前接收条件下被提供最大可行的舒适，而另一方面，保证在接收条件变得更坏时，识别结果的质量不会

过分被降级。

对于语音识别系统本身的响应，可安排：语音识别系统转换到更鲁棒的操作模式，其中接收质量较少影响语音识别的质量。例如，首先检测输入语音信号的话音活动检测器可以被切换到不太敏感模式，或者在语音识别系统的响应发生之前经过的等待周期可以被延长，以便语音识别系统完全不再对任何次噪声如此快速响应。作为选择或者另外，报警信号被输出给用户，所述报警信号最好是由语音识别系统输出的语音信号形式的更精心制作的告警消息，以使用户自己还可以对接收质量的恶化进行响应，并且例如移动到更寂静环境，或者当使用移动无线电设备时，通过搜索适当位置来改善移动无线连接。

两种响应方法的各种组合是可行的。例如，可以仅仅向用户发出报警信号或者可以仅仅执行自动转换到不太敏感操作模式。然而，可以随着转换到更鲁棒的操作模式，可以同时向用户输出报警信号，所述报警信号最好包含操作模式已经被转换的指示。此外，也有可能最初仅仅向用户输出报警信号，并且例如当接收质量未改善或者甚至恶化时，在给定时间周期之后，转换到不太敏感操作模式。

随后也可以将接收质量阈值定义为取决于相关应用的给定参数例如依赖于语音信号的不正确识别的可能结果而是可变的。尤其也有可能定义多个不同的接收质量阈值，并且逐步地激活不同的操作模式和/或每次依赖于所述接收质量阈值向用户输出报警信号。

在特别有利的实施例，当接收质量值再次超过接收质量阈值时，语音识别系统被自动复位到前一操作模式。这意味着：语音识别系统例如自动地再次变得更为敏感，以便只要用户为了响应报警信号而已经成功地创建更好的接收质量，或者当接收质量因其它原因而再次变得较好时，该语音识别系统对用户更舒适地响应。

在特别有利的实施例，只要接收质量降低到接收质量阈值以下，就断开语音识别系统的所谓插嘴（barge-in）操作模式。这意味着，语音识别系统从其中用户对语音识别系统的语音输出具有插嘴（所谓的提示）可能性的操作模式切换到“替换”操作模式，在“替换”模式中，在用户可以对语音识别系统的提示做出反应之前，用户必需首先等待提示的完整输出。此类插嘴语音识别系统一方面与用户的“替换”系统相比提供优点，即它们实质上更自然地响应，即类似于“真

实”谈话者。另一方面，有经验的用户可以更快地操作这样的语音识别系统，因为用户通常已经知道待输出的提示，所以用户可以通过过早的响应来中断提示的输出，因而实际上减少了整个对话的持续时间。

接收质量值可以以各种方式来确定。

利用话音活动检测器来确定接收质量值是特别简单的，一般来说，所述语音活动检测器已经用在语音识别系统的输入端上，用于首先检测所有的输入语音信号，并将其传送给实际的语音识别装置。随后，例如根据在用户的语音暂停中接收的背景信号，可以确定接收质量值。这意味着，例如在语音暂停期间，在输入端上测量噪声电平或者基本信号能量，以使用作接收质量的度量。例如，因而可以确定用户是位于寂静环境中还是位于喧闹环境中。此外，还可以根据对于识别结果获得的信用值或者根据其它参数，利用实际的语音识别装置自身来确定识别质量值，所述其它参数依赖于例如识别结果的质量或者为了识别而进行的努力。

然后，能够以各种方式进行语音识别系统内接收质量值的实际处理。

另一方面，话音活动检测器可以把接收质量值本身应用于例如对话控制装置。在此情况下，例如，遵从相关应用配置对话控制装置的相关应用的程序设计者可以根据接收质量阈值指定语音识别系统应当以什么方式进行反应或者什么提示将输出给用户。

作为选择或者另外地，当接收质量值落到接收质量阈值以下时，话音活动检测器还有可能向对话控制装置和/或系统的其它部件输出接收恶化指示信号。随后，对话控制装置可以相应地响应接收恶化指示信号，并向用户输出例如适当提示。而且，操作模式可以经由对话控制装置和/或其它系统部件来改变。

在特别舒适的系统中，根据引起接收质量值低于接收质量阈值的干扰的类型，更详细地分析输入信号。在此情况下，以包含相关信息的提示的形式向用户提供报警信号。这意味着，执行例如有关是否涉及用户的通信终端与语音识别系统之间的不良连接的分析，或者执行有关用户附近的背景噪声是否太响的分析。当用户接收到相应信息时，对于用户来说将更容易作出正确反应，以改善接收质量。尤其可

以利用话音活动检测器和/或实际的语音识别装置来执行该分析。

为了执行根据本发明的方法，本发明的语音识别系统一方面应当包括用于确定相应接收质量值或者噪声值的适当质量控制装置。另一方面，该语音识别系统应当包括比较器，用于将接收质量值与给定的接收质量阈值进行比较，或者用于将噪声值与给定的噪声阈值进行比较。最后，此类型的语音识别系统需要适当的控制装置，例如转换装置或合适编程的对话控制装置，用于在接收质量值降低到给定接收质量阈值以下时，或者在噪声值超过噪声阈值时，将语音识别系统转换到对噪声不太敏感和/或向用户输出报警信号的操作模式。比较器还可以被集成在该系统的其它部件中，例如被集成在话音活动检测器中或者被集成在对话控制装置中。

根据本发明的语音识别系统最好借助于适当软件在一个计算机上或在设备的语音控制中高度实施。例如可以全部以软件模块的形式实现语音识别装置和对话控制装置。借助于适当软件还可以实现用于生成适当提示的装置，例如所谓的 TTS 转换器（文本-语音转换器）。作为选择，还有可能使用提示播放器，为用户重放给定的预先记录的提示。当然，该系统必需包括用于语音输入和语音输出的设施；这样的设施将以硬件形式来实现，例如，麦克风和扬声器或用于连接到通信网的接口，利用终端经由该接口可以访问该语音识别系统。

特别地，语音识别系统的各个部件能够位于经由适当网络彼此连接的不同位置中，而不是在单个设备内例如在单个服务器上实现语音识别系统。根据本发明的语音识别系统可以与所有类型的应用结合使用。特别地，它还可同时用于多个应用，也就是，作为例如时间共享系统形式的服务系统，以便它构成用于各种应用的相应用户接口。

附图说明

下面将根据两个实施例并结合附图详细说明本发明。其中：

图 1 显示根据本发明的语音识别系统的第一实施例的方框图；和
图 2 显示根据本发明的语音识别系统的第二实施例的方框图。

具体实施方式

两个附图中所示的实施例都是所谓的插嘴型的语音识别系统 1。

麦克风 2 连接到每个语音识别系统 1, 以便输入语音信号 S_i 。为了给用户输出声音信号或者语音输出(提示), 相应的扬声器 3 连接到语音识别系统 1。作为选择, 麦克风 2 和扬声器 3 还可以存在于远离语音识别系统 1 并且经由合适的接口例如电话网等连接到语音识别系统 1 的终端中。

来自两个语音识别系统 1 的麦克风 2 的输入信号首先到达回声滤波器 4 (还称之为回声消除器)。在用户说话的情况下, 即使语音识别系统 1 自身仍然在输出提示 S_0 , 该回声滤波器 4 也用于除去提示 S_0 的回声, 该回声自输入信号以衰减形式被叠加到语音信号 S_i 上。

语音信号 S_i 随后被施加到话音活动检测器 5 上, 该检测器 5 测量信号能量, 并因而测量用户的话音活动。该话音活动检测器 5 一方面用于检测语音信号的开始和结束以及用户发声中的语音暂停。另一方面, 该话音检测器输出插嘴信号 S_{BI} , 该信号 S_{BI} 被施加到文本-语音转换器 8 (以下称之为 TTS 转换器 8) 和对话控制装置 10, 以便在用户与仍然在输出的提示相撞的情况下, 阻止 TTS 转换器 8 进一步输出提示。

在此情况下, 插嘴转换单元 9 被插入在用于话音活动检测器 5 的插嘴信号 S_{BI} 的输出引线中。系统 1 实际上专门在插嘴转换单元 9 处于闭合状态时适合于插嘴。然而, 当转换器打开时(图 1 所示的状态), 语音识别系统 1 在提示输出期间对用户的语音输入不作出反应。

语音信号 S_i 从话音活动检测器 5 施加到自动语音识别装置 7, 在装置 7 中执行对语音信号 S_i 的实际分析, 以识别其中包含的语音信息。识别结果随后以文本格式或者以机器可读格式被施加给对话控制装置 10。

最后, 对话控制装置 10 以预定方式(也就是, 根据给定的对话协议)响应识别结果, 并且例如从存储器 12 中选择用户预期的信息或者在其中存储用户输入的数据。而且, 对话控制装置驱动 TTS 转换器 8 并向该转换器提供将输出给用户的提示的数据。这些提示随后从 TTS 转换器 8 经由回声滤波器 4 施加给扬声器 3, 经由该扬声器输出这些提示。对话控制装置 10 的精确结构一般来说取决于相关应用。

根据本发明, 除了插嘴转换单元 9 之外, 附图所示的两个实施例还包括构成话音活动检测器 5 的一部分的质量控制装置 6。在所述质量控制装置 6 中确定接收质量值 S_0 。该信号可以是例如在发声之前和/或

在用户的语音暂停中由话音活动检测器 5 测量的背景噪声电平的倒数

值。

接收质量值 S_0 被施加给对话控制装置 10, 在比较器 11 比较接收质量值 S_0 与预定接收质量阈值, 该比较器 11 例如以软件模块形式来实现。只要接收质量值 S_0 低于预定阈值, 对话控制装置 10 就把插嘴停用 (deactivation) 信号 S_{BA} 提供给插嘴转换单元 9, 其中转换器被设置到图 1 所示的位置。这意味着: 语音识别系统 1 从其中允许插嘴的操作模式转换到其中不允许插嘴的操作模式。这提供了这样的优点: 在提示输出期间在背景噪声阈值过高的情况下, 语音识别系统 1 将不会被连续中断, 因为背景噪声不与用户相关联并且不应导致提示中断。这意味着: 语音识别装置 1 转换到必然导致略微不太自然并因而对于用户来说不太舒适类型的对话类型的操作模式, 但是该操作模式更鲁棒并且实际上更适合于在相关时刻出现的接收条件。

此外, 只要接收质量值 S_0 落到预定接收质量阈值以下, 对话控制单元 10 确保 TTS 生成器 8 生成并输出相关的报警信号, 该报警信号的形式为适合的报警提示 S_W 的形式, 例如, 类型为“背景噪声太大。请在应答之前等待提示输出的完成”的消息。

对话控制装置 10 被以这样的方式进行安排: 当接收质量值 S_0 再次增加到超过预定接收质量阈值时, 自动取消插嘴停用信号 S_{BA} , 由此再次闭合插嘴转换单元 9 的转换器。

原则上, 由对话控制装置 10 的程序设计员来定义停用插嘴状态和/或向用户发出报警信号的任意接收质量阈值, 以便在需要时用户可以尽可能地查找较寂静环境。

图 2 显示根据本发明的语音识别系统 1 的一个略微不同的实施例。

在该实施例中, 在质量控制装置 6 中直接将接收质量值 S_0 与接收质量阈值进行比较; 这意味着: 比较器 11 被集成在质量控制装置 6 中。当质量控制值落到预定接收质量阈值以下时, 质量控制装置 6 输出同时停用插嘴转换单元 9 的接收恶化指示信号 S_{BB} , 并将其施加给对话控制装置 10, 以确保: 取决于程序设计, 开始经由 TTS 转换器 8 输出适当的报警提示 S_W 。

请再次注意, 图中所示的语音识别系统 1 仅是本发明的特定实施

例，并且本领域熟练技术人员将能够以各种方式修改语音识别系统 1 或者用于操作语音识别系统 1 的方法而不背离本发明的范围。

例如，对于质量控制单元 6 构成对话控制装置 10 的一部分是特别有可能的。此外，插嘴转换装置 9 还能够构成话音活动检测器 5 或对话控制装置 10 的一部分，以及被构成为例如纯软件转换器。此外，质量控制单元 6 还可以被构成为独立的纯软件模块。而且，语音识别系统可以包括附加部件或者其它功能；例如，它可以执行智能的基于语法的插嘴方法。在此情况下，插嘴信号不是由话音活动检测器生成，而是仅由语音识别装置生成，该语音识别装置首先检查输入信号是否实际上具有有意义的并因而被认为是插嘴事件的内容。进一步的变化可以包括例如说话者标识或者语音识别系统的其它精心制作。

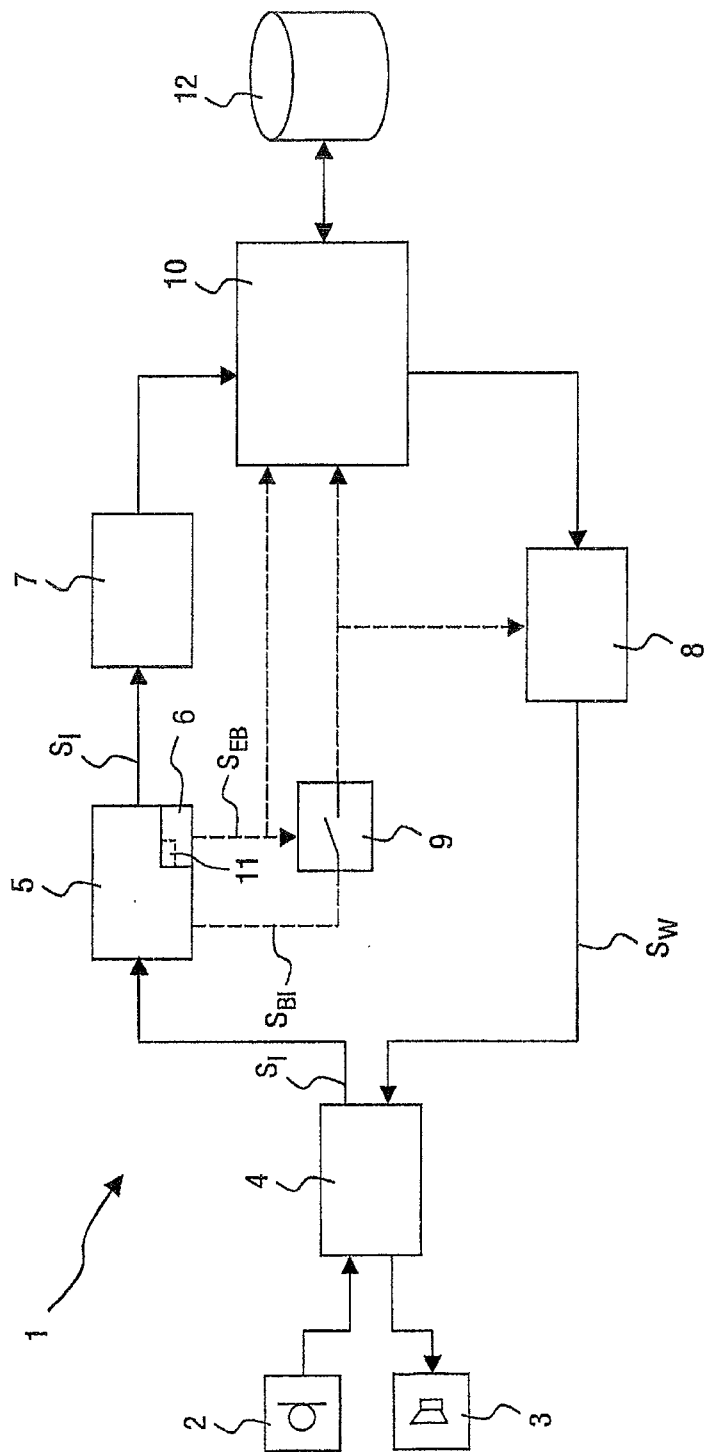


图 2