[19] 中华人民共和国国家知识产权局

[51] Int. Cl[7]
G10L 15/18

[12] 发明专利说明书

专利号 ZL 00816566.1

[45] 授权公告日 2005 年 10 月 26 日

[11] 授权公告号 CN 1224954C

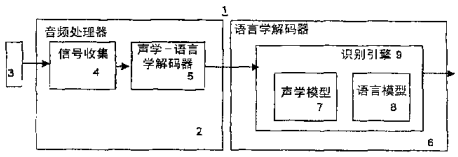
[22] 申请日 2000.11.29 [21] 申请号 00816566.1
[30] 优先权
[32] 1999.12.2 [33] FR [31] 99/15190
[86] 国际申请 PCT/FR2000/003329 2000.11.29
[87] 国际公布 WO2001/041125 法 2001.6.7
[85] 进入国家阶段日期 2002.5.31
[71] 专利权人 汤姆森许可贸易公司
地址 法国布洛里
[72] 发明人 克里斯托夫·德洛内
努尔-埃迪·塔齐尼
弗雷德里克·苏夫莱
审查员 杨 叁
[74] 专利代理机构 中科专利商标代理有限责任公司
代理人 戎志敏

权利要求书 1 页 说明书 8 页 附图 1 页

[54] 发明名称 含有固定和可变语法块的语言模型的语音识别装置

[57] 摘要

本发明涉及一种语音识别装置(1)，包括一个声音处理器(2)，用于收集音频信号，和一个语音学解码器(6)，用于确定相应于该音频信号的词顺序。本发明装置的语言学解码器包括一个语言模型(8)，它是在第一组块和第二组块的基础上确定的，第一组块至少是一个仅仅由语法确定的句法块，第二组块至少是一个由下列元素的一个，或这些元素的组合确定的句法块：一种语法，一组短语，一个 n-语法网络。



ISSN1008-4274

1．一种语音识别装置（1），包括音频处理器（2），用于收集音频信号，语言解码器（6），用于确定相应于音频信号的词顺序，解码器包括语言模型（8），其特征在于语言模型（8）是由第一组块和第二组块确定，第一组块至少是一个固定的句法块，第二组块至少是一个可变的句法块，其中，至少第一组的一个固定句法块由 BNF 型语法确定，至少第二组的一个可变句法块由一个或多个 n－语法网络确定，每个可变语法块并入 BNF 语法作为特殊符号，并包含允许从该块退出的块出口字。

2．根据权利要求 1 所述的语音识别装置，其特征在于 n－语法网络的数据由一种语法的帮助或短语表的帮助产生。

3．根据权利要求 2 所述的语音识别装置，其特征在于 n－语法网络包含相应于一个或多个下列现象的数据：简单支吾、简单重复、简单置换、想法改变、说话含糊。

含有固定和可变语法块的语言模型的语音识别装置

技术领域

本发明涉及一种语音识别装置，它包含一种语言模型，该模型是根据称为固定块和可变块的不同类语法块确定的。

背景技术

信息系统或控制系统越来越多地应用语音界面来与用户进行快速和直觉的交互作用，由于这些系统正在变得比较复杂，支持对话的方式也正变得更丰富，人们正在进入非常大量词汇的连续语音识别的领域。

已知大量词汇的连续语音识别系统的设计需要产生一个语言模型，它确定应用词汇中一个给定词以时间顺序跟在一组词中别的词后面的概率。

这种语言模型必须能再现系统用户通常使用的讲话样式：支吾、迷惑的起始、想法的改变等等。

所用语言模型的质量极大地影响语音识别的可靠性。这种质量通常是用该语言模型的因惑性指数来测量，在原理上，该指数代表选择的数目，这种选择是系统对每一被解码的词必须做的。这一指数越低，质量越高。

语言模型需要把声音信号转换成词的文本串，这是对话系统常用的步骤。然后，需要构建一种能理解的逻辑关系，使能理解口头的提问，从而作出回答。

有两种产生大词汇语言模型的标准方法：

（1）所谓的 N－语法统计模型，最常用的双语法或三语法，其要点是，假定一个词在句中的出现概率仅仅与前面的 N 个词有关，那么，它与句中的上下文无关。

考虑一个有 1000 个词汇的三语法的例子，因为它有 1000^3 个可能的

三元素组，所以，它必须确定 1000^3 的概率来定义一个语言模型，因此，需要占用相当规模的存储器和非常强的计算能力。为了解决这个问题，把词分成组，这些组由模型设计者直接确定，或者由自组织方法推导出来。

这种语言模型是由文本大全自动构造的。

（2）第二种方法的要点是借助于概率统计语法来描述语法，典型的是一种与上下文无关，依靠一组所谓 Backus Naur 公式或 BNF 公式中描述的规则来确定无上下文语法。

描述语法的规则通常是手写的，但也可以自动推导出来。在这个方法中，可参考下面的文件：

“无上下文关系的概率统计语法的基本方法”，F. Jelinek, J. D. Lafferty & R. L. Mercer, NATO ASI Series Vol. 75 pp. 345－359, 1992。

当把它们应用于自然语言系统的界面时，上面描述的模型产生了一些特殊的问题：

N－语法型语言模型（1）不能正确模拟句子中几个隔开的语法子结构之间的关系。对于句法上正确发声的句子来说，没有什么可保证在识别过程中遵守这些子结构，因此，很难确定是否该句子就是由一种或多种特殊句法结构习惯产生的这种句子或这种意义。

这些模型适合于连续的口授，但把他们应用到对话系统就有所提到的严重的缺陷。

另一方面，在 N－语法型模型中，借助于把最新实际发声的词组在一起来定义一组词，就可能考虑到支吾和重复。

基于语法（2）的模型，可以使它正确模拟句子中隔开的远程关系，也遵守特定的句法结构。对于一种给定的应用，所得到的语言的困惑常常比 N－语法型模型低。

另一方面，他们很难适应掺入有支吾、迷惑的起始等的口语型语言的描述。特别地，这些与口语型语言有关的现象不能预测，因此，似乎很难依靠其自身的特性来设计基于语法规则的语法。

此外，覆盖应用需要的规则数目很大，在没有修改这种现有规则之前，很难考虑要加入到对话中的新句子。

发明内容

本发明的目的是提供一种语音识别装置。

为实现上述目的，一种语音识别装置，包括音频处理器，用于收集音频信号，语言解码器，用于确定相应于音频信号的词顺序，解码器包括语言模型，其特征在于语言模型是由第一组块和第二组块确定，第一组块至少是一个固定的句法块，第二组块至少是一个可变的句法块，其中，至少第一组的一个固定句法块由 BNF 型语法确定，至少第二组的一个可变句法块由一个或多个 n一语法网络确定，每个可变语法块并入 BNF 语法作为特殊符号，并包含允许从该块退出的块出口字。

这两种句法块的联合，在从模拟句子元素间的依赖获得好处的同时能使有关口语语言的问题容易得到解决，这种模拟借助于一个固定句法块的帮助是容易处理的。

根据另一特征，含在第二可变块中的 n一语法网络包含允许识别下列口语现象的数据：简单支吾、简单重复、简单置换，想法改变，说话含糊。

通过确定两类实体的组合形成最终的语言模型，本发明的语言模型就能把两个系统的优点组合在一起。

固定的句法相对于某一实体保持不变，句法分析与它们相联系，而其它句法由 n一语法型网络来描述。

此外，根据改变的实施例，确定了由以前类型的一种快“触发的”自由块。

附图说明

本发明的其它特征和优点，通过非限制的特例的描述将变得更加明显，下列附图用来解释该实施例：

图 1 是语音识别系统图；

图 2 是根据本发明确定一种句法块的 OMT（直接或收发转换）图。

具体实施方式

图 1 是用于语言识别的一个实施例设备 1 的方块图。这个设备包括音频信号处理器 2，用来执行来自话筒 3 由信号收集电路 4 产生的音频信号的数字化。处理器 2 也把数字样本转换成从预先确定的字母中选择的声音符号。为此目的，它包括声学一语音学解码器 5。语言学解码器 6 处理这些符号，用来确定对一个符号顺序 A，最可能给出顺序 A 的词顺序 W。

语言学解码器使用声学模型 7 和语言模型 8，它们是基于假设搜索算法 9 实现的。例如，声学模型是所谓的“隐式 Markor”模型（或 HMM）。在本实施例中使用的语言模型是基于一种有 Backus Naur 公式的句法规则帮助说明的语法。用该语法模型为搜索算法提供假设。后者，它是合适的识别引擎，在本实施例中，是一种基于 Viferbi 型算法的搜索算法，并称为“n－最佳”。该 n－最佳型算法确定了在句子分析的每一步的 n 个最可能的词顺序。在句子的末了，从这 n 个候选中选择最可能的解决方案。

上一节中的概念本身已为业内人士所熟知，但特别与 n－最佳算法有关的信息在下面的著作中给出：

“用于语言识别统计方法”F. Jelinek, MIT Press 1999 ISBN 0－262－10066－5 pp. 79－84。其它算法也可实现。特别是“最大有效长度搜索”型算法，n－最佳算法只是它的一个例子。

本发明的语言模型使用图 2 中说明的一类或两类句法块：固定型块，可变型块。

固定句法块是根据 BNF 型句法确定的，有五种规则如下：

（a）<符号 A>＝<符号 B>|<符号 C>（或符号）

（b）<符号 A>＝<符号 B><符号 C>（和符号）

（c）<符号 A>＝<符号 B>？（选项符号）

（d）<符号 A>＝“辞典字”（辞典分配）

（e）<符号 A>＝p｛<符号 B>，<符号 C>，……<符号 X>｝（符号 B><符号 C>）（……）（符号 I><符号 J>）

（所有列举的符号的不重复置换具有这样的限制：符号 B 必须在符号 C 之前，符号 I 在符号 J 之前……）

规则（e）的实现，在法国专利申请 No. 9915083 中有详细解说，题目是“Dispsitif de reconnaissance Vocale meltant en oeuvre une regle syntaxiqne de permutation”（实现句法置换规则的语音识别装置），THOMSon Mnltimedia on November, 1999。

可变块通过与以前相同的 BNF 句法、短语表，或根据词汇表和相应的 n－语法网络，或根据把这三者联合起来确定。但是，这一信息被系统地转换到 n－语法网络中，并且，如果可变块的确定是通过一个 BNF 文件来实现的话，那么不能保证产生在句法上正确符合这一语法的唯一的句子。

可变块是由下面公式的概率 P（S）（在三语法情况下）确定的，P（S）表示 n 个词 W_i 出现字串 S 的概率。

$P(S) = \Pi_{1,n} P(Wi)$

其中 $P(W_i) = P(W_i | W_{i-1}, W_{i-2})$

对于每一个可变块，存在一个专门的块出口字，该字在 n－语法网络中表现为与通常词一样，但是它没有语音的线索并允许从该块中退出。

一旦确定了这些句法块（n－语法型或 BNF 型），它们可以再一次作作原子用于高阶结构中：

在 BNF 块中的情况下，较低水平的块可用来代替辞典的用途，以及在其它规则中使用。

在 n－语法型块的情况下，较低水平的块代替词 W_i，因此，几个块可以按照给定的概率链接起来。

一旦确定了 n －语法网络，它可与以前作为特殊符号描述的 BNF 语法相结合。多个 n－语法网络根据需要可结合在 BNF 语法中。用于 BNF 型块确定的置换在识别引擎中按布尔变量的搜索算法进行处理，在常规地实现这种类型的修剪期间，布尔变量用于指向该搜索。

可以看到，可变块出口符号也可解释成用于对上述块倒行的符号，该块本身可以是固定的或可变的。

• 触发器的配置

上述体系仍不足以描述大词汇量人/机对话应用的语言模型。根据一

个改变的实施例，增补了触发器的机构。该触发器能把某种意义给予一个词或一块，使它与某个元素相连系。例如，假定词“documentary”在音视节目的电子导视的上下文中被识别。这个词可以与一组词，如“Wildlife, Sports, tourism，等等”相联系。这些词都有与“documentary”相关的意义，其中的一个可能就是期望与它相联系的一个。

为这样做，我们将用<block>表示以前已描述过的一个块，并且用::<block>表示通过在识别算法过程中的一个瞬间完成的这一块，也就是说，在 n－最佳搜索算法中，它出现在当前被解码的链中。

例如，我们可以有：

<wish>＝I would like to go to | I want to visit.

<city>＝Lyon | Paris | London | Rennes.

<sentence>＝<wish><city>

于是:: <wish>将是：“I would like to go to”为由 Viterbj 概率算法产生的通路的那一部分：

I would like to go to Lyon

I would like to go to Paris

I would like to go to London

I would like to go to Rennes

并且将等于“I want to visit”为其它部分。因此，语言模型的触发器可定义如下：

如果<符号>::属于问题中可能实现的给定的子组，那么，另一个符号<T(符号)>，它是当前符号的目标符号，可减少为它的正常扩展域的一个子部分，也就是说，如果在解码链中没有触发器的话，减少为它的正常扩展域（简化器触发器），或者被激活并可用在从每一个属于所谓“激活器候选者”（激活器触发器）句法块中在出口分出的非零分支因子中。

注意：

没有必要描述触发过程的所有块。

符号的目标如果被用于语言模型中的多种方式，它可以是这一符号本身。

对于一个块，可以只存在它的实现部分的一个子部分，该实现部分是触发机构的一个分量，是补充而不是触发器本身。

激活器触发器的目标可以是一个可选的符号。

简化器触发机构，在我们的语言模型中，可处理话题一致的重复。关于触发器概念的附加信息可在已列出的参考文件中找到，特别在 p. 245－253 中。

激活器触发机构可以模拟高度变化词尾的语言中某种自由的句法组。

应注意到，激活器，它们的目标和关于目标的限制可以人工确定或用一个自动的处理来获得，例如用最大墒方法。

• 对口语的修正

上面描述的结构确定了语言模型的句法，没有对支吾、恢复、错误开始、想法改变，等等的修正，这些修正在一种口语式样中是需要的。与口语有关的这些现象由于它们的不可预见性很难通过一种语法来识别。n－语法网络比较适合于识别这类现象。

与口语有关的这些现象可分成五类：

简单的支吾：我想（errr……无声）去里昂。

简单的重复：其中句子的一部分（经常是限定词和冠词，但有时整块句子），十分简单地被重复：我想去（去去去）里昂。

简单的置换：在这一过程中，一种表达，沿着其方式，被同样意义的另一种表达代替，但它们的句法结构是不同的：我想访问（errv 去）里昂。

改变想法：在表达过程中，句子的一部分用不同意义的部分来校正：我想去里昂，（errr 去巴黎）。

说话含糊不清：我想去（巴黎 Errr）巴黎。

前两种现象最经常：支吾约占这些现象中的 80%。

本发明的语言模型处理这些现象如下：

简单支吾：

简单支吾用创建与在相关语言中标记支吾的语言学痕迹有关的词来处理，并且把他们当作在有关语言模型中同样的其它词一样（紧接着静

音的出现概率，等等），并在该语音模型中（同清晰度，等等）。

已经注意到，简单的支吾发生在句子专门的地方，例如在第一动词与第二个动词之间。为了对他们进行处理，根据本发明的规则的一个例子是：

<动词组>=<第一动词> <n－语法网络><第二动词>

简单重复：

简单重复通过缓冲存储器技术来处理，该存储器包含有在解码的这一步当前被分析的句子。在该语言模型中，存在有在缓冲存储器中固定的分支概率。缓冲存储器的出口以恢复缓存器激活之前达到的状态与块状语言模型相联。

实际上，缓存器包含句子当前段的最后一块。并且这一块可以重复。另一方面，如果它是倒数第二个块，它决不可能用这样一种缓存器来处理，并且整个句子必须重检查。

当包含与冠词有关的重复时，对于有关的句子，缓存器借助于改变数和性包括该冠词和它的有关的形式。

例如在法国，对“ele”的缓存包含“du”和“des”。事实上、性和数的修改是很经常的。

简单置换和想法改变：

简单置换用创建有关块组来处理，在这些块之间，简单置换是可能的，也就是说，存在从某块退出和分支到该组的一个其它块起始的可能性。

对于简单的置换，块退出与同一组内，支持同样意义的块触发相耦合。

对于想法改变，或者没有触发，或者触发支持不同意义的块。

不对触发再分表，而对支吾用后验分析分类是可能的。

说话含糊：

还可作为简单重复来处理。

处理支吾这种模型的优点（除了简单支吾之外）是关联组的建立，考虑到语义学信息冗余的存在，提高了关于无支吾句子的识别率。另一方面，计算的负担较重。

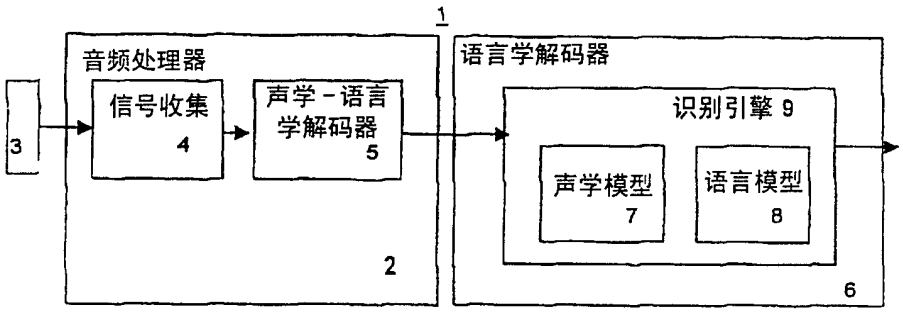

1
音频处理器
信号收集
4
声学－语言
学解码器
5
3
2
语言学解码器
识别引擎 9
声学模型
7
语言模型
8
6

图 1

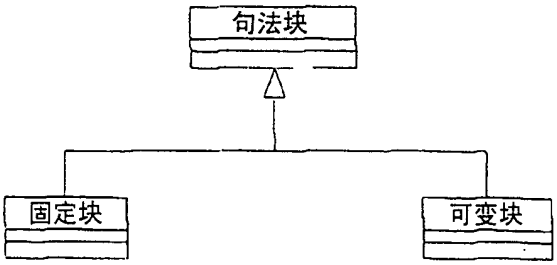

句法块
固定块
可变块

图 2