

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G10L 15/10

G10L 15/22



[12] 发明专利说明书

[21] ZL 专利号 00803588.1

[45] 授权公告日 2004 年 12 月 1 日

[11] 授权公告号 CN 1178203C

[22] 申请日 2000.2.4 [21] 申请号 00803588.1

[30] 优先权

[32] 1999. 2. 8 [33] US [31] 09/248,513

[86] 国际申请 PCT/US2000/002903 2000.2.4

[87] 国际公布 WO2000/046791 英 2000.8.10

[85] 进入国家阶段日期 2001.8.8

[71] 专利权人 高通股份有限公司

地址 美国加利福尼亚州

[72] 发明人 毕宁 张承纯 H·加鲁达德瑞

A·P·德雅科

审查员 潘宁媛

[74] 专利代理机构 上海专利商标事务所

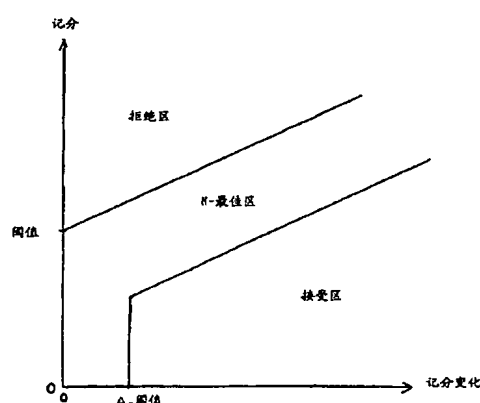
代理人 洪玲

权利要求书 4 页 说明书 7 页 附图 2 页

[54] 发明名称 语音识别拒绝方法

[57] 摘要

一种捕获发音的语音识别拒绝方法，包括接受发音、对发音应用 N-best 算法或拒绝发音。该方法包括将所述发音与第一存储单词进行比较以产生第一记分；将所述发音与第二存储单词进行比较以产生第二记分；确定所述第一和第二记分之间的差值；确定所述第一记分与该差值的比率；根据所述比率处理所述发音。如果所述第一记分与该差值的比率在第一值域范围内，就接受所述发音；如果所述第一记分与该差值的比率在第二值域范围内，就对所述发音应用 N-best 算法；如果所述第一记分与该差值的比率在第三值域范围内，则拒绝所述发音。第一、第二与第三值域范围确定了第一存储单词和差值之间的线性关系。



ISSN 1008-4274

权 利 要 求 书

1. 一种在话音识别系统中捕获发音的方法，包括以下步骤：
将所述发音与第一存储单词进行比较以产生第一记分；
将所述发音与第二存储单词进行比较以产生第二记分；
确定所述第一和第二记分之间的差值；
确定所述第一记分与该差值的比率；
根据所述比率处理所述发音。
2. 如权利要求 1 所述的方法，其特征在于，进一步包括：
如果所述第一记分与该差值的比率在第一值域范围内，就接受所述发音；
如果所述第一记分与该差值的比率在第二值域范围内，就对所述发音应用 N-best 算法；
如果所述第一记分与该差值的比率在第三值域范围内，则拒绝所述发音。
3. 如权利要求 1 所述的方法，其特征在于，所述差值对应于第一记分和第二记分之间的记分变化。
4. 如权利要求 1 所述的方法，其特征在于，所述第一存储单词包括话音识别系统的词汇表中的最佳候选单词，而所述第二存储单词包括话音识别系统的词汇表中的下一最佳候选单词。
5. 如权利要求 1 所述的方法，其特征在于，所述第一记分包括最接近的比较结果，所述第二记分包括下一最接近的比较结果。
6. 如权利要求 1 所述的方法，其特征在于，所述第一存储单词和所述第二存储单词包括线性预测编码系数。
7. 如权利要求 1 所述的方法，其特征在于，所述第一存储单词和所述第二存储单词包括倒谱系数。
8. 如权利要求 1 所述的方法，其特征在于，所述第一存储单词和所述第二存储单词包括带通滤波器输出。
9. 如权利要求 2 所述的方法，其特征在于，所述第一、第二与第三值域范围确定了第一存储单词和差值之间的线性关系。
10. 如权利要求 1 所述的方法，其特征在于，所述差值包括最接近的比较结果与下一最接近的比较结果之间的差值。
11. 一种话音识别系统，其特征在于包括：

声学处理器，配置成从一发音的数字化语音样本中提取语音参数；以及
耦合到所述声学处理器的处理器，该处理器配置成：将所述发音与第一存储单词进行比较以产生第一记分；将所述发音与第二存储单词进行比较以产生第二记分；确定所述第一和第二记分之间的差值；确定所述第一记分与该差值的比率；根据所述比率处理所述发音。

12. 如权利要求 11 所述的话音识别系统，其特征在于，所述处理器进一步用于：

如果所述第一记分与该差值的比率在第一值域范围内，就接受所述发音；

如果所述第一记分与该差值的比率在第二值域范围内，就对所述发音应用 N-best 算法；

如果所述第一记分与该差值的比率在第三值域范围内，则拒绝所述发音。

13. 如权利要求 11 所述的话音识别系统，其特征在于，所述差值对应于第一记分和第二记分之间的记分变化。

14. 如权利要求 11 所述的话音识别系统，其特征在于，所述第一存储单词包括话音识别系统的词汇表中的最佳候选单词，而所述第二存储单词包括话音识别系统的词汇表中的下一最佳候选单词。

15. 如权利要求 11 所述的话音识别系统，其特征在于，所述第一记分包括最接近的比较结果，所述第二记分包括下一最接近的比较结果。

16. 如权利要求 11 所述的话音识别系统，其特征在于，所述第一存储单词和所述第二存储单词包括线性预测编码系数。

17. 如权利要求 11 所述的话音识别系统，其特征在于，所述第一存储单词和所述第二存储单词包括倒谱系数。

18. 如权利要求 11 所述的话音识别系统，其特征在于，所述第一存储单词和所述第二存储单词包括带通滤波器输出。

19. 如权利要求 12 所述的话音识别系统，其特征在于，所述第一、第二与第三值域范围确定了第一存储单词和差值之间的线性关系。

20. 如权利要求 11 所述的话音识别系统，其特征在于，所述差值包括最接近的比较结果与下一最接近的比较结果之间的差值。

21. 一种话音识别系统，其特征在于包括：

模式比较逻辑，用于将所述发音与第一存储单词进行比较以产生第一记分，以及将所述发音与第二存储单词进行比较以产生第二记分；

判断逻辑，用于确定所述第一和第二记分之间的差值；确定所述第一记分与该差值的比率；并根据所述比率处理所述发音。

22. 如权利要求 21 所述的话音识别系统，其特征在于，所述判断逻辑进一步用于：

如果所述第一记分与该差值的比率在第一值域范围内，就用于接受所述发音；

如果所述第一记分与该差值的比率在第二值域范围内，就对所述发音应用 N-best 算法；

如果所述第一记分与该差值的比率在第三值域范围内，则用于拒绝所述发音。

23. 如权利要求 21 所述的话音识别系统，其特征在于，所述差值对应于第一记分和第二记分之间的记分变化。

24. 如权利要求 21 所述的话音识别系统，其特征在于，所述第一存储单词包括话音识别系统的词汇表中的最佳候选单词，而所述第二存储单词包括话音识别系统的词汇表中的下一最佳候选单词。

25. 如权利要求 21 所述的话音识别系统，其特征在于，所述第一记分包括最接近的比较结果，所述第二记分包括下一最接近的比较结果。

26. 如权利要求 21 所述的话音识别系统，其特征在于，所述第一存储单词和所述第二存储单词包括线性预测编码系数。

27. 如权利要求 21 所述的话音识别系统，其特征在于，所述第一存储单词和所述第二存储单词包括倒谱系数。

28. 如权利要求 21 所述的话音识别系统，其特征在于，所述第一存储单词和所述第二存储单词包括带通滤波器输出。

29. 如权利要求 22 所述的话音识别系统，其特征在于，所述第一、第二与第三值域范围确定了第一存储单词和差值之间的线性关系。

30. 如权利要求 21 所述的话音识别系统，其特征在于，所述差值包括最接近的比较结果与下一最接近的比较结果之间的差值。

31. 如权利要求 21 所述的话音识别系统，其特征在于还包括：

声学处理器，用于从一发音的数字化语音样本中提取语音参数；以及

32. 如权利要求 31 所述的话音识别系统，其特征在于，所述判断逻辑进一步用于：

如果所述第一记分与该差值的比率在第一值域范围内，就接受所述发音；

如果所述第一记分与该差值的比率在第二值域范围内，就对所述发音应用 N-best 算法；

如果所述第一记分与该差值的比率在第三值域范围内，则拒绝所述发音。

33. 如权利要求 31 所述的话音识别系统，其特征在于，所述差值对应于第一记分和第二记分之间的记分变化。

34. 如权利要求 31 所述的话音识别系统，其特征在于，所述第一存储单词包括话音识别系统的词汇表中的最佳候选单词，而所述第二存储单词包括话音识别系统的词汇表中的下一最佳候选单词。

35. 如权利要求 31 所述的话音识别系统，其特征在于，所述第一记分包括最接近的比较结果，所述第二记分包括下一最接近的比较结果。

36. 如权利要求 31 所述的话音识别系统，其特征在于，所述第一存储单词和所述第二存储单词包括线性预测编码系数。

37. 如权利要求 31 所述的话音识别系统，其特征在于，所述第一存储单词和所述第二存储单词包括倒谱系数。

38. 如权利要求 31 所述的话音识别系统，其特征在于，所述第一存储单词和所述第二存储单词包括带通滤波器输出。

39. 如权利要求 32 所述的话音识别系统，其特征在于，所述第一、第二与第三值域范围确定了第一存储单词和差值之间的线性关系。

40. 如权利要求 31 所述的话音识别系统，其特征在于，所述差值包括最接近的比较结果与下一最接近的比较结果之间的差值。

说明书

话音识别拒绝方法

技术领域

本发明一般涉及通信领域，尤其涉及话音(voice)识别系统。

背景技术

话音识别(VR)代表赋予机器模拟智能以识别用户命令或用户有声命令并有助于人们与机器联系的最重要的技术之一。VR 还代表理解人们语音(speech)的关键技术。应用根据声学语音信号而恢复语言消息的技术的系统称为话音识别器。话音识别器一般包括声学处理器与字解码器，前者提取实现输入原始语音的 VR 所必需的信息承载(bearing)特征或矢量序列，后者对此特征或矢量序列作解码以产生与输入发音(utterance)相对应的有意义且想要的输出格式，诸如一语言文字序列。为提高某一给定系统的性能，需要训练该系统具备诸有效参数。换言之，系统在能最佳地工作之前需要进行学习。

声学处理器代表话音识别器中的前端语音分析子系统。响应于输入语音信号，声学处理器表征随时间变化的语音信号的适当表示。声学处理器应丢弃无关的信息，如背景噪声、信道失真、讲话者特征和说话方式等。有效的声学处理法向话音识别器提供增强的声学鉴别能力。为此，要分析的有用特性就是短时谱包络。表征短时谱包络的两种常用谱分析技术是线性预测编码(LPC)与基于滤波器组(filter bank)的谱建模。在转让给本发明受让人的 5,414,796 号美国专利(这里引用其全部内容作为参考)以及 L. B. Rabiner 与 R. W. Schafer 所著的“语音信号的数字处理(Digital Processing of Speech Signals)”(396-453(1978))(这里也引用其全部内容作为参考)中，都描述了示例的 LPC 技术。

由于安全性，VR(常称作语音识别)的应用变得越发重要。例如，VR 可以用来代替人工按压无线电话键板上的按钮。这在用户驾驶车辆的同时启动电话呼叫时尤其重要。在使用无 VR 的电话时，驾驶者必须从方向盘上腾出一只手，看着电话键板，同时按下按钮拨号呼叫。这些动作易发生车祸。语音使能电话(即为语音识别所设计的电话)能让驾驶者一面安排电话呼叫，一面继续观察路面。一种免提的车辆配套系统还能让驾驶者在启动呼叫时仍可保持将双手放在方向盘上。

语音识别装置分成讲话者相关装置或讲话者无关装置。讲话者无关装置能接受来自任何用户的话音命令。讲话者相关装置更常用，它被训练成识别来自特定用户的命令。讲话者相关 VR 装置通常工作于两个阶段，即训练阶段与识别阶段。在训练阶段中，VR 系统提醒用户把系统词汇中的每个单词讲一二遍，让系统能认识用户对这些特定单词或短语的语音特征。或者，对于语音 VR 装置，通过读出专门编写的一篇或多篇短文从而覆盖这种语言中的全部音素来实现训练。免提车辆配套系统的一种示例词汇表可以包括键板上的数字；关键词“call(通话)”、“send(发送)”、“dial(拨号)”、“cancel(取消)”、“clear(清除)”、“add(添加)”、“delete(删除)”、“history(历史)”、“program(编程)”、“yes(是)”与“no(否)”；以及预定数目的经常通话的同事、朋友或家人的名字。训练一旦完成，用户说出训练过的关键词就能在识别阶段启动呼叫。例如，如果名字“John”是训练过的名字之一，则用户说出短语“Call John(呼叫 John)”就能启动对 John 的呼叫。VR 系统将识别单词“Call”和“John”，并拨出用户先前作为 John 的电话号码所输入的号码。

VR 系统的通过量可以定义为用户成功通过识别任务的实例的百分数。一项识别任务通常包括多个步骤。例如，在用无线电话进行话音拨号时，通过量指用户以 VR 系统成功地完成电话呼叫的次数的平均百分数。以 VR 实现成功电话呼叫所需的步骤数依据各次呼叫而变化。一般而言，VR 系统的通过量主要取决于两个因素，即 VR 系统的识别精度与人机接口。用户主观感受到的 VR 系统性能以通过量为基础。因此，为提高通过量，要求 VR 系统具有高的识别精度和智能型人机接口。

发明内容

本发明旨在一种具有高识别精度和智能型人机接口以提高通过量的 VR 系统。因此，在本发明的一个方面中，在语音识别系统中俘获发音的方法最好包括以下步骤：将所述发音与第一存储单词进行比较以产生第一记分；将所述发音与第二存储单词进行比较以产生第二记分；确定所述第一和第二记分之间的差值；确定所述第一记分与该差值的比率；根据所述比率处理所述发音。

在本发明的第二方面中，语音识别系统最好包括配置成从一发音的数字化语音样本中提取语音参数的声学处理器；以及与该声学处理器耦合的处理器，该处理器配置成：将所述发音与第一存储单词进行比较以产生第一记分；将所述发音与第二存储单词进行比较以产生第二记分；确定所述第一和第二记分之间的差值；

确定所述第一记分与该差值的比率；根据所述比率处理所述发音。

在本发明的再一个方面中，话音识别系统最好包括：一装置，用于将所述发音与第一存储单词进行比较以产生第一记分；一装置，用于将所述发音与第二存储单词进行比较以产生第二记分；一装置，用于确定所述第一和第二记分之间的差值；一装置，用于确定所述第一记分与该差值的比率；一装置，用于根据所述比率处理所述发音。

在本发明的还有一个方面中，话音识别系统最好包括从一发音的数字化语音样本中提取语音参数的装置；以及一装置，用于将所述发音与第一存储单词进行比较以产生第一记分；将所述发音与第二存储单词进行比较以产生第二记分；确定所述第一和第二记分之间的差值；确定所述第一记分与该差值的比率；根据所述比率处理所述发音。

附图概述

图 1 是话音识别系统的框图。

图 2 是 VR 系统拒绝方案的记分(score)对记分变化的曲线，示出了拒绝、N-best 和接受区域。

较佳实施例的详细描述

根据图 1 所示的一个实施例，话音识别系统 10 包括模拟-数字转换器(A/D) 12、声学处理器 14、VR 模板数据库 16、模式比较逻辑 18 与判断逻辑 20。VR 系统 10 可以驻留在例如无线电话或免提车辆配套系统中。

在 VR 系统 10 处于语音识别阶段时，某人(未图示)说出某个单词或短语，产生一语音信号。该语音信号由常规换能器(也未图示)转换成语音电信号 $s(t)$ 并提供给 A/D 12，A/D 12 按公知的采样方法(如脉冲编码调制(PCM))将该语音信号 $s(t)$ 转换成数字化的语音样本 $s(n)$ 。

将语音样本 $s(n)$ 供给声学处理器 14 作参数确定。声学处理器 14 产生一组提取的参数，以对输入语音信号 $s(t)$ 的特征进行建模。可根据许多公知的语音参数确定技术中的任一种确定这些参数，这些技术包括例如上述 5,414,796 号美国专利中所述，进行编码并应用基于快速富利叶变换(FFT)的倒谱系数的语音编码器。声学处理器 14 可实现为数字信号处理器(DSP)。DSP 可包括语音编码器。或者，可把声学处理器 14 实现为语音编码器。

也可在 VR 系统 10 的训练期间进行参数确定，其中把 VR 系统 10 的所有词汇表单词的一组模板按路由传给 VR 模板数据库 16 以永久地存储在其中。可以有利地把 VR 模板数据库 16 实现为任何常规形式的非易失性存储媒体，诸如快闪存储器。这使得在断开 VR 系统 10 的电源时，让模板保持在 VR 模板数据库 16 中。

将该组参数提供给模式比较逻辑 18。模式比较逻辑 18 有利地检测一发音的起点与终点，计算动态声学特征(诸如，时间导数、二阶时间导数等)，通过选择相关帧而压缩声学特征，并且对静态与动态声学特征进行量化。如在 Lawrence Rabiner 与 Biing-Hwang Juang 著的“语音识别基础(Fundamentals of Speech Recognition)”(1993)中，对终点检测、动态声学特征求导、模式压缩与模式量化的各种公知方法作了描述，这里引用其全部内容作为参考。

模式比较逻辑 18 将该组参数与存储在 VR 模板数据库 16 中的所有模板作比较。把比较结果或该组参数与存储在 VR 模板数据库 16 中的所有模板之间的距离提供给判断逻辑 20。判断逻辑 20 可以：(1)从 VR 模板数据库 16 中选出与该组参数最紧密匹配的模板，或可以：(2)应用“N-best”选择算法，该算法在预定匹配阈值内选择 N 个最紧密的匹配；或可以：(3)拒绝该组参数。若使用 N-best 算法，则询问此人想要哪一种选择。判断逻辑 20 的输出是对已说过词汇表中的哪一个单词的判断。例如，在 N-best 的情况下，此人可能说“John Anders”，而 VR 系统 10 可能回答“Did you say John Andrews(你说了 John Anders 是吗)？”。然后，此人再答“John Anders”。接着，VR 系统 10 可能再答“Did you say John Anders(你说了 John Auders 是吗)？”。此人再答“Yes(是)”，此时 VR 系统 10 就启动一电话呼叫的拨号。

可以有利地把模式比较逻辑 18 与判断逻辑 20 实现为微处理器。或者, 可把模式比较逻辑 18 与判断逻辑 20 实现为任何常规形式的处理器、控制器或状态机。VR 系统 10 可以是例如专用集成电路(ASIC)。VR 系统 10 的识别精度可以是 VR 系统 10 能多好地正确识别说出的词汇表中的单词或短语的量度。例如, 95% 的识别精度表明, VR 系统 10 在 100 次中有 95 次正确地识别出词汇表中的单词。

在一实施例中，如图 2 所示，将记分与记分变化曲线分成接受、N-Best 与拒绝诸区域。这些区域按公知的线性鉴别分析技术用线来分离，这些技术在 Richard O. Duda 与 Peter E. Hart 著的“模式分类与情景分析 (Pattern Classification and Scene Analysis)” (1973) 中有描述，这里引用其全部内容作为参考。如上所述，模式比较逻辑 18 给至 VR 系统 10 的每个发音输入分配对应于存储在 VR 模板数据

库 16 中的每一模板的比较结果或与其的距离。这些距离或“记分”有利的是在多个帧上相加的 N 维矢量空间中矢量之间的欧几里得距离。在一实施例中，矢量空间是一二十四维矢量空间，在二十个帧上累积记分，而且该记分是一整数距离。本领域的技术人员应该理解，完全可把记分表示为小数或其它值。本领域内的技术人员还应理解，可用其他度量来替代欧几里得距离，因而记分可以是例如概率量度、似然性量度等。

对于一给定的发音与一来自 VR 模板数据库 16 的给定 VR 模板，记分越低(即发音与 VR 模板之间的距离越小)，则发音与 VR 模板之间的匹配越紧密。对于每个发音，判断逻辑 20 相对于该记分同有关 VR 模板数据库 16 中第二最紧密匹配的记分(即第二最低记分)之间的差值分析有关 VR 模板数据库 16 中最紧密匹配的记分。如图 2 的曲线所示，“记分”是针对“记分变化”绘制的，且限定了三个区域。拒绝区代表记分相对高且该记分与下一最低记分之差相对小的区域。如果一发音落在拒绝区内，则判断逻辑 20 就拒绝该发音。接受区代表记分相对低且该记分与下一最低记分之差相对大的区域。如果一发音落在接受区内，则判断逻辑 20 就接受该发音。N-best 区位于拒绝区与接收区之间。N-best 区代表记分小于拒绝区的记分或该记分与下一最低记分之差大于拒绝区中的记分差值的区域。N-best 区也代表记分大于接受区中的记分或者该记分与下一最低记分之差小于接受区中的记分差值的区域，只要 N-best 区中该记分的差值大于一预定的记分变化阈值。如上所述，如果一发音落在 N-best 区内，则判断逻辑 20 就对该发音应用 N-best 算法。

在参考图 2 描述的实施例中，第一线段将拒绝区与 N-best 区分离。第一线段与“记分”轴在预定记分阈值处相交。第一线段的斜率也是预定的。第二线段将 N-best 区与接受区分离。第二线段的斜率预定为与第一线段的斜率相同，因而第一与第二线段平行。第三线段从“记分变化”轴上的一个预定变化阈值开始垂直延伸，与第二线段的终点相遇。本领域的技术人员可理解，第一与第二线段不一定要平行，可以有任意分配的斜率。另外不一定使用第三线段。

在一实施例中，记分阈值是 375，变化阈值为 28，若第二线段的终点延伸，则第二线段应在值 250 处与“记分”轴相交，因而第一与第二线段的斜率都为 1。如果记分值大于记分变化值加 375，则该发音被拒绝。反之，如果记分值大于记分变化值加 250 或该记分变化值小于 28，则对发音应用 N-best 算法。否则，接受该发音。

在参照图 2 描述的实施例中，线性鉴别分析应用了两个量纲。量纲“记分”

代表一给定的发音与从多个带通滤波器(未图示)的输出中导出的一给定 VR 模板之间的距离。量纲“记分变化”代表最低记分(即最紧密匹配的记分)与下一最低记分(即下一最紧密匹配发音的记分)之差。在另一实施例中,量纲“记分”代表一给定发音与从该发音的倒谱系数中导出的一给定 VR 模板之间的距离。在另一实施例中,量纲“记分”代表一给定发音与从该发音的线性预测编码(LPC)系数中导出的一给定 VR 模板之间的距离。从在上述 5,414,796 号美国专利中描述了导出一发音的 LPC 系数与倒谱系数的技术。

在其他实施例中,线性鉴别分析不限于二个量纲。相应地,相互对应地分析基于带通滤波器输出的第一记分、基于倒谱系数的第二记分和记分变化。或者,相互对应地分析基于带通滤波器输出的第一记分、基于倒谱系数的第二记分、基于 LPC 系数的第三记分和记分变化。本领域的技术人员很容易理解,量纲“记分”不必限于任何特定数。本领域内的技术人员容易理解,记分量纲的数目不仅仅由 VR 系统词汇表中的单词数来限定。本领域内的技术人员还应理解,所采用的记分的类型不必局限于任何特定的记分类型,而可包括本领域公知的任何记分方法。另外,本领域的技术人员也容易理解,“记分变化”的量纲数不必局限于一或任何特定数目。例如,在一实施例中,相对于最紧密匹配同下一最紧密匹配之间的记分变化来分析记分,相对于最紧密匹配同第三最紧密匹配之间的记分变化来分析记分。本领域内的技术人员很容易理解,记分变化量纲的数目不仅仅由 VR 系统词汇表中的单词数目来限定。

这样,已描述了一种以线性鉴别分析为基础的新型改进话音识别拒绝方案。本领域的技术人员应该理解,可用数字信号处理器(DSP)、专用集成电路(ASIC)、分立的门或晶体管逻辑、分立的硬件元件(诸如寄存器与 FIFO)、执行一组固件指令的处理器或任何常规可编程软件模块与处理器来实施或执行结合这里所揭示的实施例描述的各种示例的逻辑组件与算法步骤。处理器最好是微处理器,但在另一替代方式中,处理器可以是任何常规的处理器、控制器、微控制器或状态机。软件模块可驻留在 RAM 存储器、快闪存储器、寄存器或本领域中公知的任何其它形式的可写存储媒体中。技术人员还应理解,用电压、电流、电磁波、磁场或粒子、光场或粒子或其任意组合来表示以上描述中可参照的数据、指令、命令、信息、信号、位、码元和码片。

已经示出和描述了本发明的较佳实施例。然而,本领域的技术人员将明白,可对这里揭示的诸实施例作出各种替换而不违背本发明的精神或范围,因而本发

03.11.23

明只限于下述的权利要求书。

说明书附图

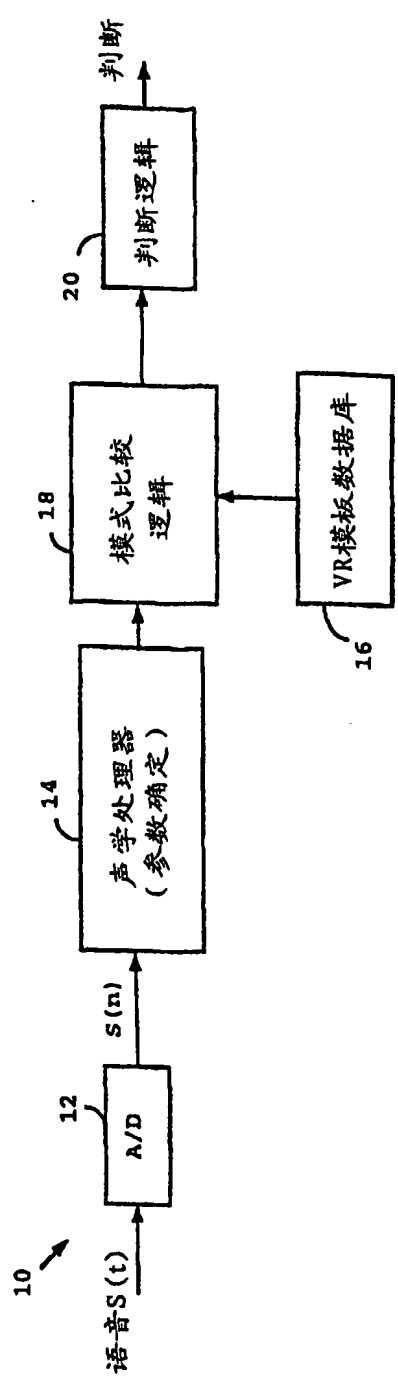


图 1

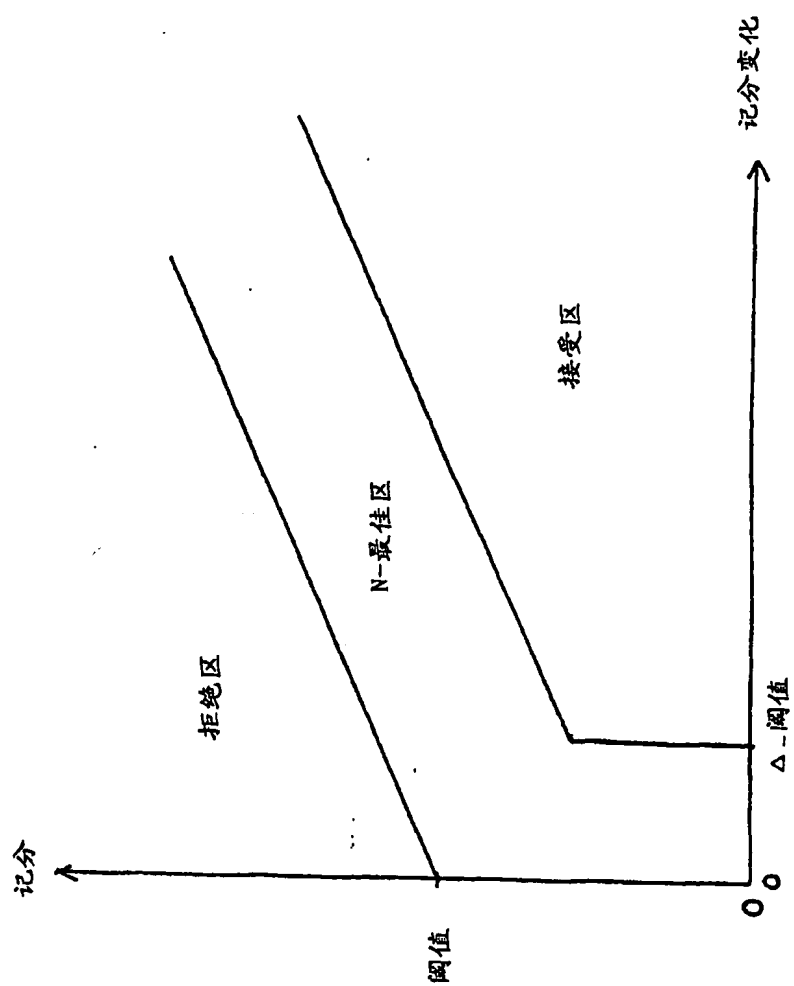


图 2