

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷
G10L 15/02



[12] 发明专利说明书

[21] ZL 专利号 00803546.6

[45] 授权公告日 2004 年 8 月 4 日

[11] 授权公告号 CN 1160698C

[22] 申请日 2000.2.8 [21] 申请号 00803546.6

[30] 优先权

[32] 1999.2.8 [33] US [31] 09/246,414

[86] 国际申请 PCT/US2000/003260 2000.2.8

[87] 国际公布 WO2000/046790 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 A·P·德雅科

审查员 胡 婧

[74] 专利代理机构 上海专利商标事务所

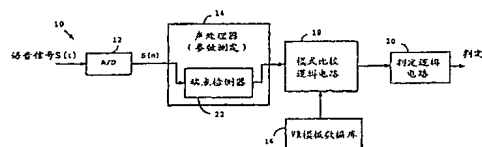
代理人 李 玲

权利要求书 2 页 说明书 10 页 附图 6 页

[54] 发明名称 噪声信号中语音的端点定位

[57] 摘要

一种用来在存在噪音时对语音进行精确的端点定位的设备,包括一处理器和一软件模块。该处理器执行软件模块的指令,比较发音与第一信噪比(SNR)门限值以确定该发音的第一起点和第一终点。然后该处理器比较时间上早于第一起点的那部分发音与第二 SNR 门限值以确定该发音的第二起点。该处理器然后还比较时间上后于第一终点的那部分发音与第二 SNR 门限值以确定该发音的第二终点。周期性地重新计算第一和第二 SNR 门限值以反映变化的 SNR 条件。第一 SNR 门限值有益地超过第二 SNR 门限值。



ISSN 1008-4274

1、一种用来检测发音在接收信号的诸帧内的端点的设备，其特征在于它包含：

一处理器，以及

一由该处理器执行的软件模块，比较发音与第一门限值以确定该发音的第一起点和第一终点，比较时间上早于第一起点的那部分发音与第二门限值以确定该发音的第二起点以及比较时间上后于第一终点的那部分发音与第二门限值以确定该发音的第二终点，其中第一和第二门限值从所述发音的信噪比每帧计算一次。

2、如权利要求1所述的设备，其中，第一和第二门限值基于信噪比。

3、如权利要求1所述的设备，其中，周期性地重新计算第一和第二门限值。

4、如权利要求1所述的设备，其中，第一门限值超过第二门限值。

5、如权利要求1所述的设备，其中，第二终点和第二起点间的差值受预定的最大和最小长度界限约束。

6、一种检测发音在接收信号的诸帧内的端点的方法，其特征在于它包含以下步骤：

比较发音与第一门限值以确定该发音的第一起点和第一终点；

比较时间早于第一起点的那部分发音与第二门限值以确定该发音的第二起点；以及

比较时间上后于第一终点的那部分发音与第二门限值以确定该发音的第二终点，其中第一和第二门限值从所述发音的信噪比每帧计算一次。

7、如权利要求6所述的方法，其中，第一和第二门限值基于信噪比。

8、如权利要求6所述的方法，还包含周期性地重新计算第一和第二门限值的步骤。

9、如权利要求6所述的方法，其中，第一门限值超出第二门限值。

10、如权利要求6所述的方法，还包含由预定的最大和最小长度界限约束第二终点和第二起点间差值的步骤。

11、一种用来检测发音在接收信号的诸帧内的端点的设备，其特征在于它

包含：

用来比较发音与第一门限值以确定该发音的第一起点和第一终点的装置；

用来比较时间上早于第一起点的那部分发音与第二门限值以确定该发音的第二起点的装置；以及

用来比较时间上后于第一终点的那部分发音与第二门限值以确定该发音的第二终点的装置，其中第一和第二门限值从所述发音的信噪比每帧计算一次。

12、如权利要求 11 所述的设备，其中，第一和第二门限值基于信噪比。

13、如权利要求 11 所述的设备，还包含用来周期性地重新计算第一和第二门限值的装置。

14、如权利要求 11 所述的设备，其中，第一门限值超出第二门限值。

15、如权利要求 11 所述的设备，还包含由预定的最大和最小长度界限约束第二终点和第二起点间差值的装置。

噪声信号中语音的端点定位

发明背景

I、发明领域

本发明总的说来涉及通信领域，更确切地说涉及存在噪声时的语音的端点定位。

II、背景

话音识别(VR)代表赋予一机器以识别用户或用户话音命令并使人机接口便利的仿真智能的最重要技术之一。VR 还代表一种主要的人类语音理解技术。使用把声的语音信号恢复为语言消息的技术的系统称为话音识别器。话音识别器典型地包含声处理器和字译码器，前者提取话音识别来向原始语音所必需的一系列带信息的特征或矢量，后者给该系列特征或矢量译码以产生诸如对应于输入发音的一系列语言字的富有意义和所需输出格式。为提高指定系统的性能，需要训练以用有效的参数装备该系统。换言之，该系统在能发挥最佳功能前需学习。

声处理器代表话音识别器中的前端语音分析子系统。作为对输入语音信号的响应，声处理器提供适当的显示法以表征该随时间变化的语音信号。声处理器应删除诸如背景噪声、信道失真，说话者特征和说话方式的不相关信息，有效的声处理使话音识别器具有增强的声鉴别力，为达到此目的，要分析的一个有用特征是短时间频谱包络。用来表征短时间频谱包络的两个常用频谱分析技术是线性预测编码(LPC)和基于滤波器组的频谱模拟。在专利号为 5414796 的美国专利及 L.B.Rabiner 和 R.W.Schfer 所著的 1978 年版的《语音信号的数字处理》一书第 396 至第 453 页中描述了示例的 LPC 技术，前者已转让给本发明的受让人并在此完全引述供参考，后者也在此完全引述供参考。

因为安全的原因，VR(通常称为语音识别)的使用变得越来越重要。例如，VR 可用来取代在无线电话键盘上揿按钮的手工任务。这在用户驾车启动电话呼叫时特别重要。使用没有 VR 的话机时，驾驶员必须从方向盘上移出一只手，

并在揿按钮拨打该呼叫时看着电话键盘。这些动作增加了车祸的可能性。能够使用语音的话机(即设计成语音识别的话机)将让驾驶员在继续注视路面的同时发出电话呼叫。免提汽车配套话机系统将附加地准许驾驶员在呼叫发起期间保持双手在方向盘上。

语音识别设备分为依赖说话人型和不依赖说话人型两类。不依赖说话人型设备能接受任何用户的话音命令。依赖说话人型设备较常见,训练成识别特定用户的命令。依赖说话人型 VR 设备典型地工作在两个阶段:训练阶段和识别阶段。在训练阶段,VR 系统提示用户对系统词汇中的每个单词说一遍或两遍,这样系统可以了解用户说这些单词或短语的语音特征。另一个可选择的是,对于语言 VR 设备,训练是通过特别按原稿读出覆盖该语言中所有音素的一篇或更多篇短文完成的。由于免提汽车配套话机系统的示例性词汇可以包括键盘上的数字;关键词“呼叫”、“发送”、“拨打”、“取消”、“消除”、“增加”、“历史”、“程序”、“是”和“否”;以及预定数量的常叫的同事、朋友或家庭成员的名字。一旦训练完成,用户在识别阶段能够通过说这些所训练的关键词发起呼叫。例如,假设“约翰”这个名字是所训练的名字之一。用户可通过说“呼叫约翰”这个短语发起向约翰的一个呼叫。VR 系统会识别“呼叫”和“约翰”这两个单词,并拨打用户事先作为约翰的电话号码输入的那个号码。

为精确捕获发话发音以作识别,能够使用语音的产品典型地使用端点检测器确定发音的起点和终点。在传统的 VR 设备中,端点检测器依据单一信噪比(SNR)门限确定发音的端点。在两篇关于语音和声频处理的 IEEE 会刊中描述了这样的传统 VR 设备,它们是 Jean-ClaudeJunqua 等人于 1994 年 7 月发表的《一种存在噪音时用作单词界限检测的稳健算法》和 1998 年 3 月的《TIA/EIA 过度性标准 IS-733》中的 2-35 至 2-50。然而,如果 SNR 门限定得太低,VR 设备对背噪音变得太敏感,这会触发端点检测器,从而导致错误识别。反之,如果门限定得太高,则 VR 设备变得易于在发音的始点和终点丢失弱辅音。这样,需要一种存在背音噪声时使用多个自适应 SNR 门限以精确检测语音的端点的 VR 设备。

发明概要

本发明关于一种存在背景噪声时使用多个自适应 SNR 门限精确检测语音的端点的 VR 设备。因此,在本发明的一个方面,一种用来检测发音端点的设备有益地包括一处理器;以及一可由该处理器执行的软件模块,用于比较发音与第一门限值以确定该发音的第一起点和第一终点、比较时间上早于第一起点的那部分发音与第二门限值以确定该发音的第二起点、以及比较时间上后于第一终点的那部分发音与第二门限值以确定该发音的第二终点。

在本发明的另一个方面,一种检测发音的端点的方法有益地包括以下步骤:比较发音与第一门限值以确定该发音的第二起点和第一终点;比较时间上早于第一起点的那部分发音与第二门限值以确定该发音的第二起点;以及比较时间上后于第一终点的那部分发音与第二门限值以确定该发音的第二终点。

在本发明的另一方面,一种用来检测发音的端点的设备有益地包括:用来比较发音与第一门限值以确定该发音的第一起点和第一终点的装置;用来比较时间上早于第一起点的那部分发音与第二门限值以确定该发音的第二起点的装置;以及用来比较时间上后于第一终点的那部分发音与第二门限值以确定该发音的第二终点的装置。

附图简述

图 1 是话音识别系统的方框图。

图 2 是说明诸如图 1 的系统的语音识别系统检测发音的端点所执行的方法步骤的流程图。

图 3 是对于不同频带,发音的信号振幅和第一、第二自适应 SNR 门限与时间的曲线图。

图 4 是说明诸如图 1 的系统的语音识别系统的比较瞬时 SNR 与自适应 SNR 门限所执行的方法步骤的流程图。

图 5 是无线话机中语音端点检测器的瞬时信噪比(分贝)与信噪比估算值(分贝)的曲线图。

图 6 是免提汽车配套话机系统中语音端点检测器的瞬时信噪比(分贝)比信噪比估算值(分贝)的曲线图。

较佳实施例的详细描述

依照一实施例,如图1所示,话音识别系统10包括模数转换器(A/D)12、声处理器14、VR模板数据库16、模式比较逻辑电路18和判定逻辑电路20。声处理器14包括端点检测器22。VR系统10可驻留在,例如无线话机或免提汽车配套话机系统中。

当VR系统10在语音识别阶段时,某人(未示出)说一单词或短语,产生一语音信号。该语音信号用普通变换器(亦未示出)转换为电语音信号 $s(t)$,将其提供给A/D转换器,A/D转换器按照已知的采样方法(如脉冲编码调制(PCM))将其转换为数字化语音采样 $s(n)$ 。

语音采样 $s(n)$ 提供给声处理器14进行参数测定。声处理器14产生模拟输入语音信号 $s(t)$ 的特征的一组参数。参数可依照许多已知语言参数测定技术中的任何一个来确定,例如包括语音编码器编码及如前面提到的专利号为5414796的美国专利中描述的使用基于快速富里叶变换(FFT)的倒频谱系数。声处理器14可实施成数字信号处理器(DSP)。该DSP可包括语音编码器。另外,声处理器14可实施成语音编码器。

在VR系统10的训练期间也执行参数确定,其中,VR系统10所有词汇的一组模板传送给VR模板数据库16,永久存储在其中。数据库16可有益地实施成诸如闪存的任何一种形式的非易失性存储媒介。这使得在关掉VR系统10的电源时这些模板能保留在VR模板数据库16中。

该组参数提供给模式比较逻辑电路18。模式比较逻辑电路18有益地检测发音的起点和终点,计算动态声特征(诸如时间导数、二次时间导数等),通过选择相关帧而压缩声特征,量化静态和动态声特征。例如,Lawrence Rabiner和Biing-Hwang Juang所著的《语音识别基本原理》(1993年版)中描述了端点检测、动态声特征推导、模式压缩及模式量化的各种已知方法,在此完全引述供参考。模式比较逻辑电路18比较该组参数和存储在VR模板数据库16中的所有模板。该组参数与存储在VR模板数据库16中的所有模板的比较结果或距离提供给判定逻辑电路20。判定逻辑电路20从VR模板数据库16选择最匹配该组参数的模板。在另一种方案中,判定逻辑电路20可使用传统的“N-best”选择算法,“N-best”选择算法在预定的匹配门限之内选择N个最接近匹配。然后询问该人有意于哪个选择。判定逻辑电路20的输出是对说了词汇中哪个词的判定。

模式比较逻辑电路 18 和判定逻辑电路 20 可有益地实施成一微处理器。VR 系统 10 可以是例如一专用集成电路(ASIC)。VR 系统 10 的识别精确度是对 VR 系统 10 正确识别词汇中的所说单词或短语的程度的测试。例如, 95% 的识别精确度表明 VR 系统 10 每 100 次中有 95 次正确识别词汇中的单词。

声处理器 14 中的端点检测器 22 确定属于每个语音发音的起点和终点的参数。端点检测器 22 用来捕获有效发音, 有效发音或者用作语音训练阶段中的一语音模板或者在语音识别阶段与语音模板比较以发现最好的匹配。在存在背景噪声时端点检测器 22 减少 VR 系统 10 的错误, 从而提高诸如对无线话机的话音拨打和话音控制的功能的稳健性。如下面参照图 2 所作的详细描述, 在端点检测器 22 中确定了两个自适应信噪比门限以捕获有效发音。第一门限高于第二门限。第一门限用来发现发音中诸如辅音的相对弱的部分。可适当调谐这两个自适应 SNR 门限以使 VR 系统 10 或者对噪声稳健或者对任何语言部分敏感。

在一实施例中第二门限是诸如上述专利号为 5414796 美国专利中描述声码器的 13k 比特每秒(kbps)声码器中的半速率门限, 第一门限比 13kbps 声码器中的全速率大 4 至 10 分贝。这两个门限有益地自适应背景 SNR, 背景 SNR 可每 10 或 20 毫秒地进行估算。这是需要的, 因为背景噪声(即公路噪声)在汽车中是变化的。在一实施例中, VR 系统 10 驻留在无线电话手机的声码器中, 端点检测器 22 在 0.3-2kHz 和 2-4kHz 这两个频带中计算 SNR。在另一实施例中, VR 系统 10 驻留在免提汽车配套话机系统中, 端点检测器 22 在 0.3-2kHz、2-3kHz 和 3-4kHz 这三个频带中计算 SNR。

依照一实施例, 端点检测器执行图 2 的流程图中所示的方法步骤, 以检测发音的端点。图 2 中示出的算法步骤可有益地用传统的数字信号处理技术实施。

在步骤 100 中清除数据缓冲器和称为 GAP 的参数。将用 LENGTH 表示的参数设定为等于称作 HEADER_LENGTH 的参数。称作 LENGTH 的参数跟踪端点正被检测的发音的长度。各种参数可有益地存储在端点检测器中的寄存中。数据缓冲器可有益地是循环缓冲器, 如果无人说话则循环缓冲器节省存储空间。声处理器(未示出)包括端点检测器, 它以每个发音固定数目帧实时处理语音发音。在一实施例中, 每帧有十毫秒。端点检测器必须从起点“回顾”某

一数目的语音帧, 因为声处理器(未示出)执行实时的处理。HEADER 的长度确定从起点回顾多少帧, HEADER 的长度可以是例如 10 至 20 帧。完成步骤 100 后, 算法进入步骤 102。

如以下参照图 4 所描述, 在步骤 102 中加载一帧语音数据并更新或重新计算 SNR 估算值。这样, SNR 估算值按每帧更新以自适应改变 SNR 条件。如下面参照图 4 和图 6 所描述, 计算第一和第二 SNR 门限。第一 SNR 门限高于第二 SNR 门限。完成步骤 102 后, 算法进入步骤 104。

在步骤 104 中当前或瞬时 SNR 与第一 SNR 门限比较。如果一预定数目 N 个连续帧的 SNR 大于第一 SNR 门限, 算法进入步骤 106。另一方面, 如果 N 个连续帧的 SNR 不比第一门限大, 算法进入步骤 108。在步骤 108 中算法用包含在 HEADER 中的帧更新数据缓冲器。然后算法回到步骤 104。在一实施例中, 数目 N 是 3。与 3 个连续帧比较是作平均之用。例如, 假如只用一个帧, 该帧包含噪音峰值。产生的 SNR 则不是在 3 个连续帧上平均的 SNR 的指示。

在步骤 106 中, 加载语音数据的下一个帧并更新 SNR 估算值。然后算法进入步骤 110。在步骤 110 中, 比较当前 SNR 与第一 SNR 门限, 以确定发音的端点。如果 SNR 小于第一 SNR 门限, 算法进入步骤 112。另一方面, 如果 SNR 不小于第一 SNR 门限, 算法进入步骤 114。在步骤 114 中清除参数 GAP 且参数 LENGTH 增加 1。然后算法返回步骤 106。

在步骤 112 中, 参数 GAP 增加 1, 然后算法进入步骤 116。在步骤 116 中参数 GAP 与称作 GAP 门限的参数比较。参数 GAP 门限表示通话期间单词间的间隙。参数 GAP 门限可有益地设定为 200 至 400 毫秒。如果 GAP 大于 GAP 门限, 算法进入步骤 118。还是在步骤 116 中, LENGTH 这个参数与称作 MAX_LENGTH 的参数比较, 下面结合步骤 154 对此进行描述, 如果长度大于或等于最大长度, 算法进入步骤 118。然而, 如果在步骤 116 中 GAP 不大于 GAP 门限, 并且 LENGTH 不大于或等于 MAX_LENGTH, 算法进入步骤 120。在步骤 120 中参数 LENGTH 增加 1。然后算法返回步骤 106 加载语音数据的下一帧。

在步骤 118 中, 算法开始向后寻找发音的起点。算法回看保存在 HEADER 中的帧, HEADER 可有益地包含 20 个帧。将称作 PRE_START 的参数设定为等于 HEADER。算法还开始寻找发音的终点, 将称作 PRE_END 的参数设定为

等于 LENGTH 减 GAP。然后算法进入步骤 122 和 124。

在步骤 122 中, 将指针 i 设定为等于 PRE_START 减 1, 清除称作 GAP_START 的参数(即设定 GAP_START 等于零)。指针 i 表示发音的起点。然后算法进入步骤 126。类似地, 在步骤 124 中, 设定指针 j 等于 PRE_END, 清除称作 GAP_END 的参数。指针 j 表示发音的终点。然后算法进入步骤 128。如图 3 所示, 在相对两端有箭头的第一行段说明发音的长度。该行的两端表示发音的真正起点和终点(即: END 减 SATRT)。在相对的两端有箭头的第二行段在第一行段下面示出, 表示 PRE_END 减 PRE_START 的值, 最左端表示指针 i 的初始值而最右端表示指针 j 的初始值。

在步骤 126 中, 算法加载帧号 i 的当前 SNR。然后算法进入步骤 130。类似地, 在步骤 128 中, 算法加载帧号 j 的当前 SNR。然后算法进入步骤 132。

在步骤 130 中, 算法把帧号 i 的当前 SNR 与第二 SNR 门限比较。如果当前 SNR 小于第二 SNR 门限, 算法进入步骤 134。另一方面, 如果当前 SNR 不小于第二 SNR 门限, 算法进入步骤 136。类似地, 在步骤 132 中, 算法把帧号 j 的当前 SNR 与第二 SNR 门限比较。如果当前 SNR 小于第二 SNR 门限, 算法进入步骤 138。另一方面, 如果当前 SNR 不小于第二 SNR 门限, 算法进入步骤 140。

在步骤 136 中, 清除 GAP_START 且指针 i 减 1。然后算法返回步骤 126。类似地, 在步骤 140 中清除 GAP_END 且指针 j 增加 1。然后算法返回步骤 128。

在步骤 134 中 GAP_START 增加 1。然后算法进入步骤 142。类似地, 在步骤 138 中 GAP_END 增加 1。然后算法进入步骤 144。

在步骤 142 中把 GAP_START 与称作 GAP_START_THRESHOLD 的参数比较。参数 GAP_START_THRESHOLD 表示说出的单词的音素间的间隙, 或快速连续交谈中相邻单词间的间隙。如果 GAP_START 大于 GAP_START_THRESHOLD, 或者如果指针 i 小于或等于零, 算法进入步骤 196。另一方面, 如果 GAP_START 不大于 GAP_START_THRESHOLD, 且指针 i 不小于或等于零, 算法进入步骤 148。类似地, 在步骤 144 中 GAP_END 与称作 GAP_END_THRESHOLD 的参数比较。参数 GAP_END_THRESHOLD 表示说出的单词的音素间的间隙, 或快速连续的交谈中相邻单词间的间隙。如果 GAP_END 大于 GAP_END_THRESHOLD, 或如果指针 j 大于或等于

LENGTH, 算法进入步骤 150。另一方面, 如果 GAP_END 不大于 GAP_END_THRESHOLD, 且指针 j 不大于或等于 LENGTH, 算法进入步骤 152。

在步骤 148 中, 指针 i 减 1。然后算法返回步骤 126。类似地, 在步骤 152 中, 指针 j 增 1。然后算法返回到步骤 128。

在步骤 146 中, 称作 START 的参数表示发音的真正起点, 设定为等于指针 i 减 GAP_START。然后算法进入步骤 154。类似地, 在步骤 150 中, 称作 END 的参数表示发音的真正终点, 设定为等于指针 j 减 GAP_END。然后算法进入步骤 154。

在步骤 154 中, 比较 END 减 START 的差值与称作 MIN_LENGTH 的参数, MIN_LENGTH 是表示比 VR 设备词汇中最短单词长度小的一长度的预定值。还将 END 减 START 的差值与参数 MAX_LENGTH 作比较, MAX_LENGTH 是表示比 VR 设备词汇中最长单词的长度还大的一长度的预定值。在一实施例中, MIN_LENGTH 是 100 毫秒, MAX_LENGTH 是 2.5 毫秒。如果 END 减 START 的差值大于或等于 MIN_LENGTH 且小于或等于 MAX_LENGTH, 则已捕获有效发音。另一方面, 如果 END 减 START 的差值小于 MIN_LENGTH 或大于 MAX_LENGTH, 该发音无效。

在图 5 中, 绘出驻留在无线话机中的端点检测器的 SNR 估算值的一组示例性第一和第二 SNR 门限。例如, 假如 SNR 估算值是 40 分贝, 则第一门限是 19 分贝而第二门限近似 8.9 分贝, 在图 6 中, 绘出驻留在免提汽车配磁话机系统中的端点检测器的 SNR 估算值(分贝)对瞬时 SNR(分贝)的曲线, 并示出基于 SNR 估算的一组示例性第一和第二 SNR 门限。例如, 假如瞬时 SNR 是 15 分贝, 则第一门限近似 15 分贝而第二门限近似 8.2 分贝。

在一实施例中, 依照图 4 的流程图中说明的步骤执行结合图 3 描述的估算步骤 102、106 和比较步骤 104、110、130 和 132。在图 4 中, 通过遵循用短划线封闭并标记为参考字 102(为简单起见)的步骤执行估算 SNR 的步骤(图 3 的步骤 102 或步骤 106)。在步骤 200 中, 用频带能量(BE)值和前一帧的平滑频带能量值(E^{SM})计算当前帧的平滑频带能量值(E^{SM})如下:

$$E^{SM} = 0.6E^{SM} + 0.4BE$$

完成步骤 200 的计算后, 执行步骤 202。在步骤 202 中, 确定当前帧的平

滑背景能量值(B^{SM})为 1.03 倍前一帧的平滑背景能量值(B^{SM})和当前帧的平滑频带能量值(E^{SM})的最小值如下:

$$B^{SM} = \min(1.03B^{SM}, E^{SM})$$

完成步骤 202 的计算后, 执行步骤 204。在步骤 204 中, 确定当前帧的平滑信号能量值(S^{SM})是 0.97 倍前一帧的平滑信号能量值(S^{SM})和当前帧的平滑频带能量值(E^{SM})的最大值如下:

$$S^{SM} = \max(0.97S^{SM}, E^{SM})$$

完成步骤 204 的计算后, 执行步骤 206。在步骤 206 中, 从当前帧的平滑信号能量值(S^{SM})和当前帧的平滑背景能量值(B^{SM})计算当前帧的 SNR 估算值(SNR^{EST})如下:

$$SNR^{EST} = 10\log_{10}(S^{SM}/B^{SM})$$

完成步骤 206 的计算后, 通过做步骤 208 的比较而执行比较瞬时 SNR 与估算 SNR(SNR^{EST})以确定第一或第二 SNR 门限的步骤(对第一 SNR 门限来说是图 3 的步骤 104 或步骤 110, 对第二 SNR 门限来说是图 3 的步骤 130 或步骤 132), 步骤 208 的比较用短划线封闭并标记为参考数字 104(为简间起见)。步骤 208 的比较使用下列等式求瞬时 SNR(SNR^{INST})。

$$SNR^{INST} = 10\log_{10}(BE/B^{SM})$$

因此, 在步骤 208 中, 依照下面的方程式比较当前帧的瞬时 SNR(SNR^{INST})与第一或第二 SNR 门限:

$$SNR^{INST} > \text{门限}(SNR^{EST})?$$

在一实施例中, VR 系统驻留于无线话机中, 可通过把当前帧的 SNR 估算值(SNR^{EST})定位在水平轴上并把第一和第二门限看作是与所示第一和第二门限曲线的相交点从图 5 的曲线图获得第一和第二 SNR 门限。在另一实施例中, VR 系统驻留于免提汽车配套话机系统中, 可通过把当前帧的 SNR 估算值(SNR^{EST})定位在水平轴上并把第一和第二门限看作是与所示第一和第二门限曲线的相交点从图 6 的曲线图获得第一和第二 SNR 门限。

可依照任何已知的方法计算瞬时 SNR(SNR^{INST}), 包括例如专利号为 5742734 和 5341456 的美国专利中描述的 SNR 计算方法, 该两专利已转让给本发明的受让人, 在此完全引述供参考。可初始化 SNR 估算值(SNR^{EST})为任何值, 但可有益地如下所述地初始化。

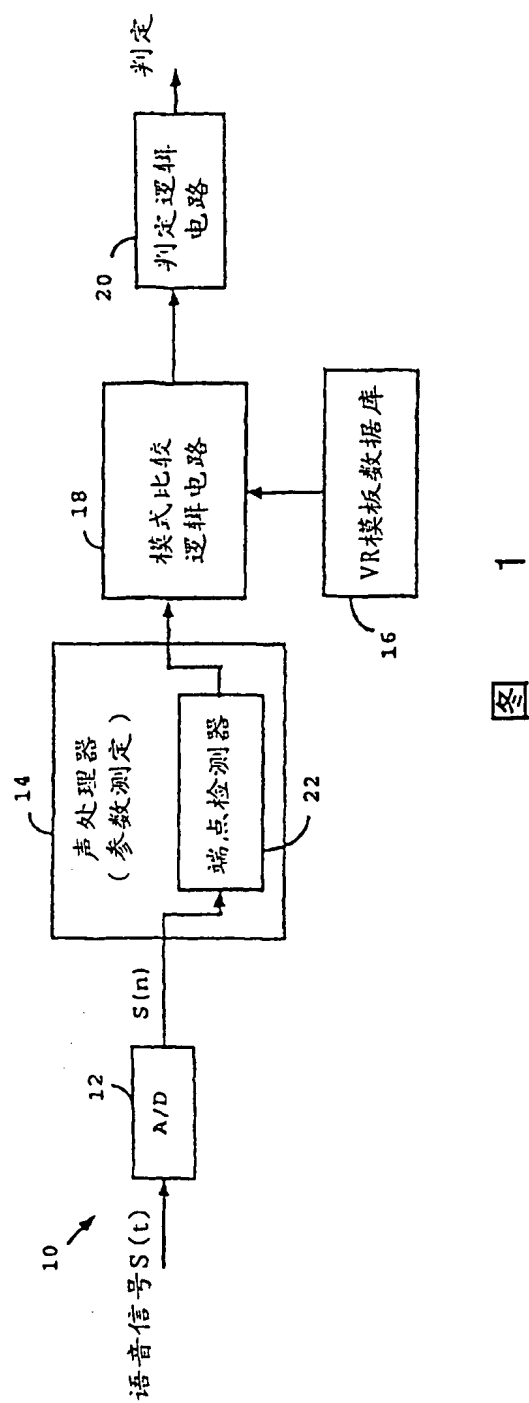
在一实施例中, VR 系统驻留于天线话机中, 设定低频带(0.3-2kHz)的平滑频带能量(E^{SM})的初始值(即第一帧中的值)等于第一帧的输入信号频带能量(BE)。还设定高频带(2 至 4kHz)平滑频带能量(E^{SM})的初始值等于第一帧的输入信号频带能量(BE)。设定平滑背景能量(B^{SM})的初始值等于低频带的 5059644 和 高频带 5059644(这些单位是信号能量的量化电平, 它是从输入信号的数字化采样的平方和计算的)。平滑信号能(S^{SM})的初始值设定为等于低频带的 3200000 和 高频带 320000。

在另一实施例中, VR 系统驻留于免提汽车配套话机系统中, 设定低频带(0.3 至 2kHz)平滑频带能量(E^{SM})的初始值(即第一帧中的值)等于第一帧的输入信号频带能量(BE)。还设定中频带(2 至 3kHz)和 高频带(3 至 4kHz)的平滑频带能量(E^{SM})的初始值等于第一帧的输入信号频带能量(BE)。设定平滑背景能(B^{SM})的初始值等于低频带的 5059644、中频带的 5059644 和 高频带 5059644。设定平滑信号能量(S^{SM})的初始值等于低频带的 3200000、中频带的 250000 和 高频带的 70000。

因此, 已描述一种在存在噪声时用于对语音进行精确端点定位的新颖、改进方法和设备。描述的实施例有益地或者通过设置适当高的第一 SNR 门限值避免误触发端点检测器, 或者通过设置适当低的第二 SNR 门限值不错过任何弱的语音部分。

本领域的熟练人员明白: 结合此处揭示的实施例描述的各种说明性逻辑块和算法步骤可实施或执行于数字信号处理器(DSP), 专用集成电路(ASIC), 分立的门电路或晶体管逻辑电路、诸如寄存器和 FIFO 的分立的硬件元件, 执行一组固件指令的处理器或任何传统的可编程软件模块及处理器。该处理可有益地是微处理器, 但作为可选项, 该处理器可以是任何传统的处理器、微控制器或状态机。该软件模块可驻留于 RAM 存储器、闪存、寄存器或任何其他形式的本领域中已知的可写存储媒介。熟练人员还会认识到: 上述描述中可参考的数据、指令、命令、信息、信号、位、码元和码片有益地用电压、电流、电磁波、磁场或磁性粒子、光场或光粒子或任何有关的组合表示。

这样已示出并描述本发明的较佳实施例。然而, 对本领域中传统的熟练人员说明显的是: 可不脱离本发明的精神或范围对此处揭示的实施例进行许多改动。因经, 将依照以下权利要求限制本发明。



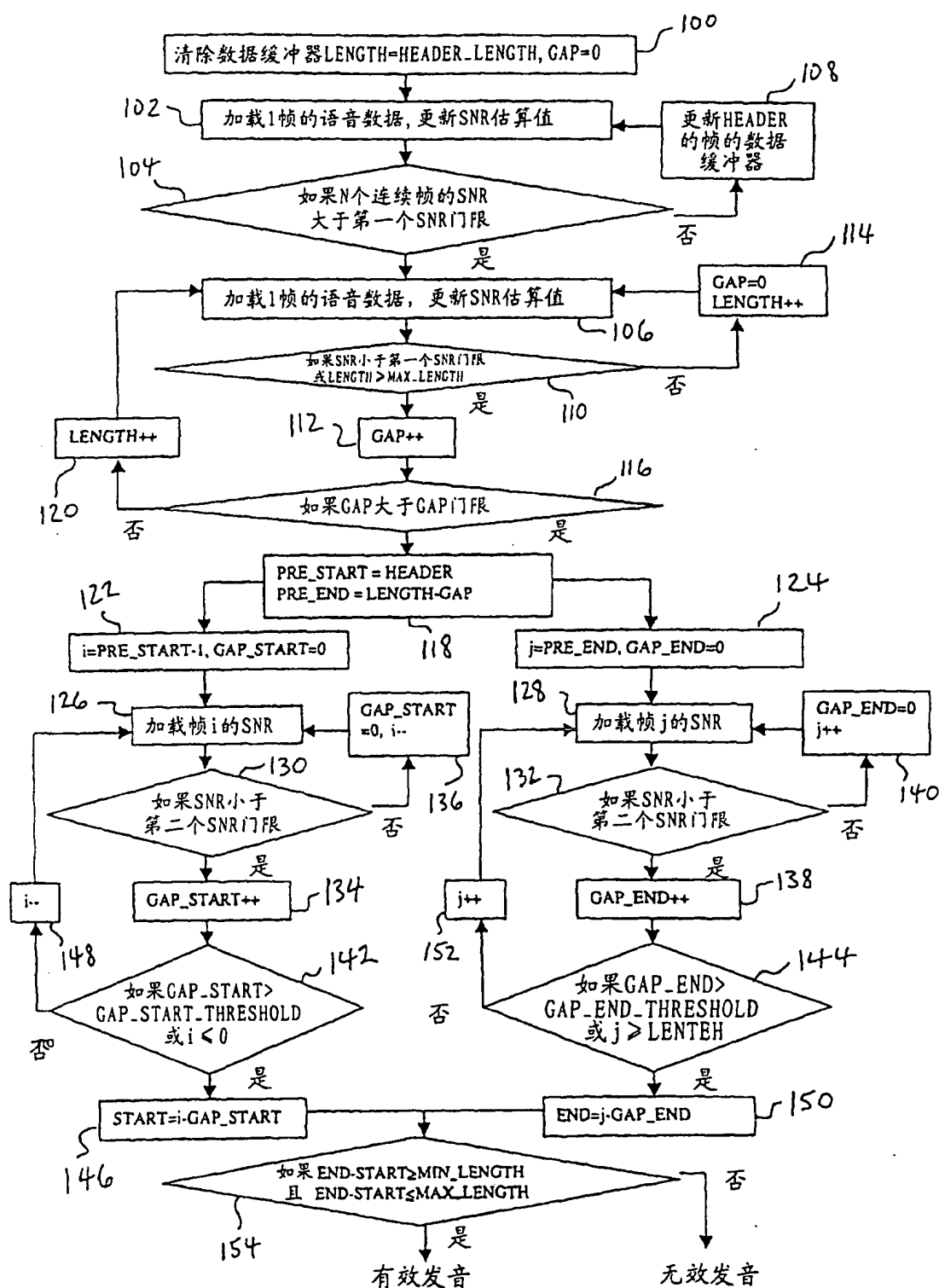


图 2

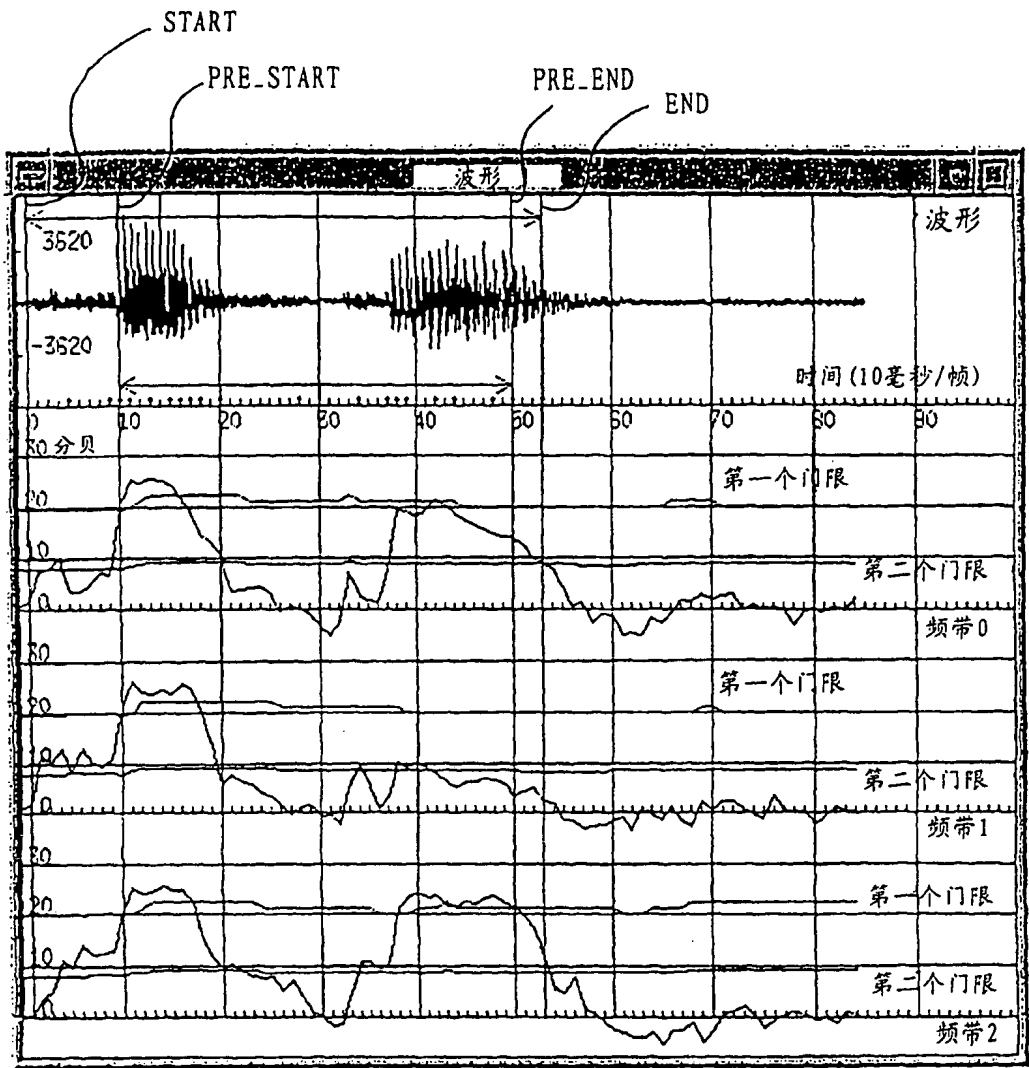


图 3

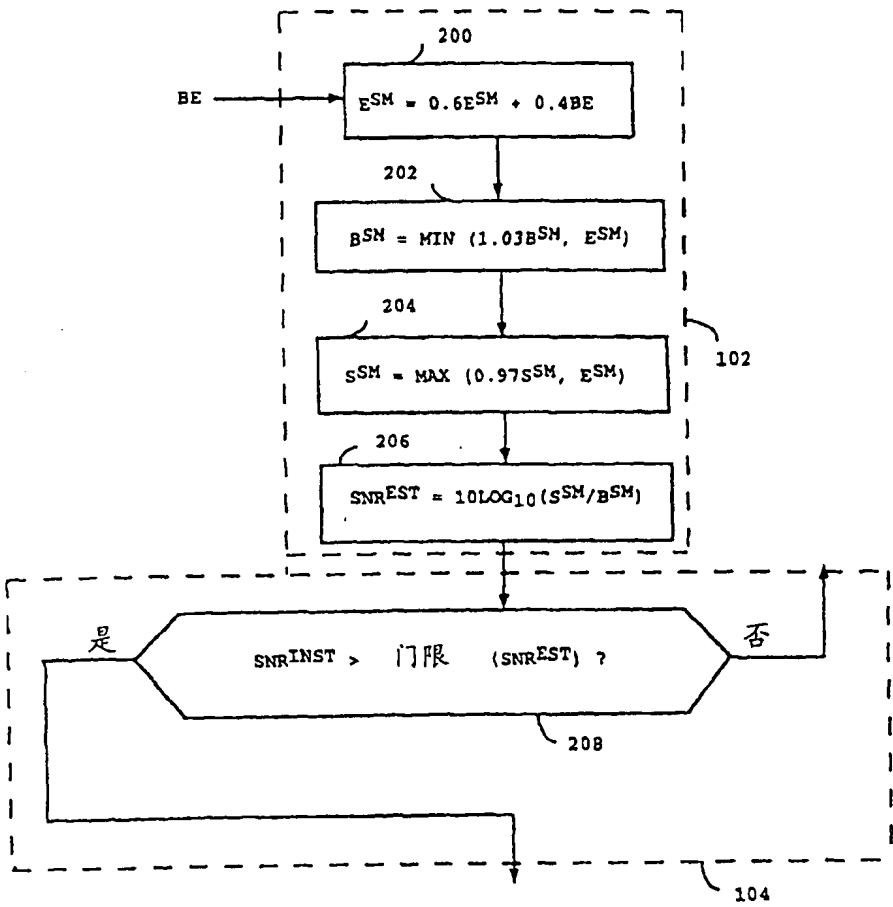


图 4

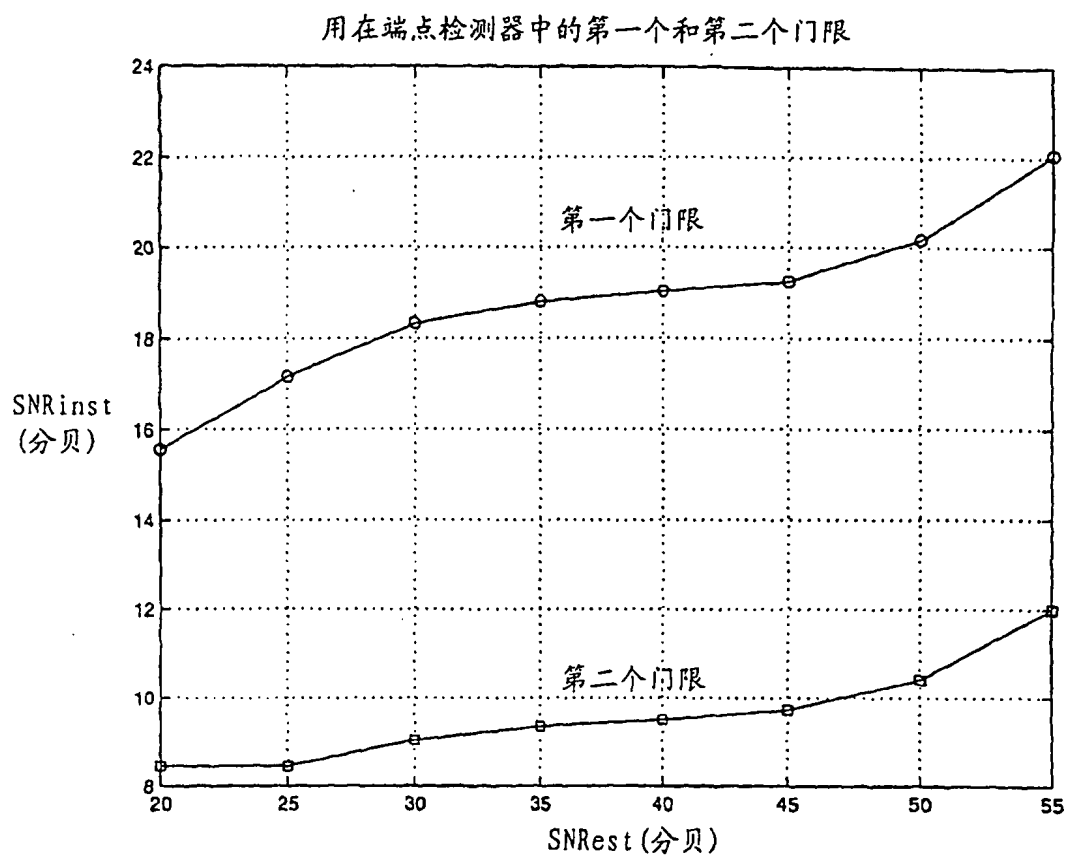


图 5

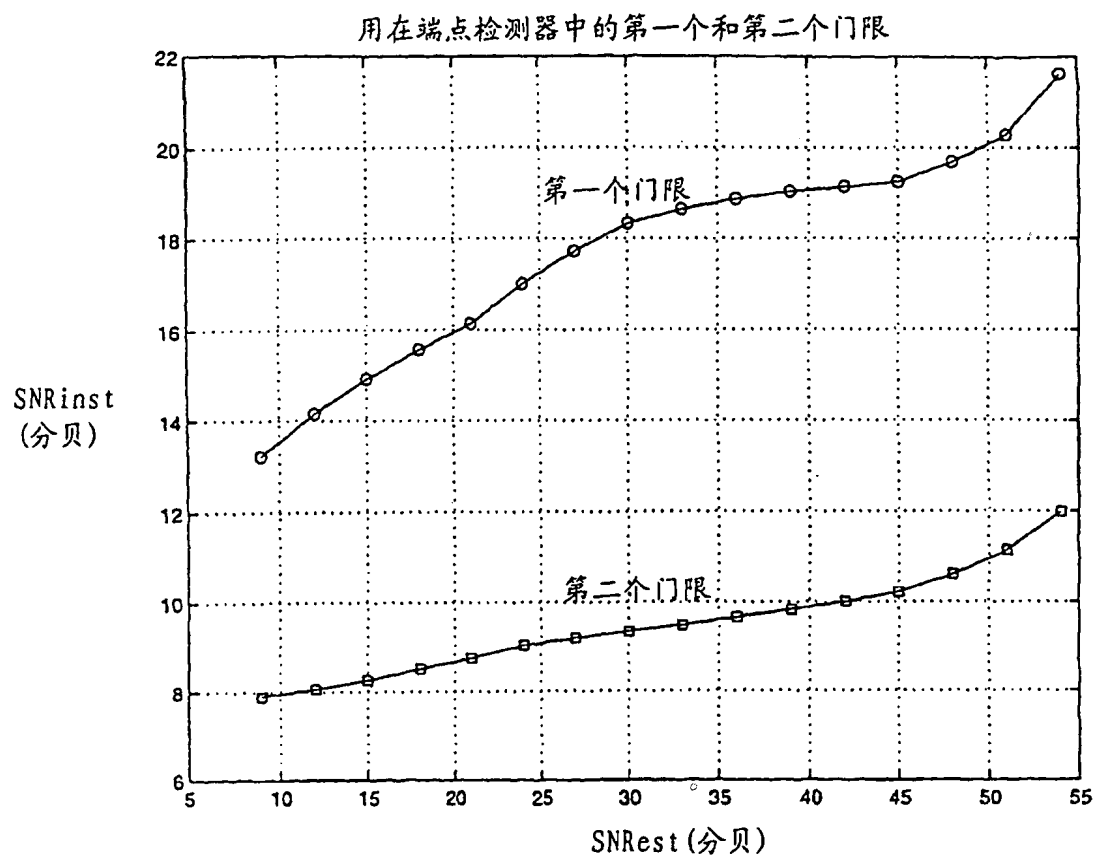


图 6