

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G10L 15/00

G10L 19/04 H04M 1/253



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 98808942.4

[45] 授权公告日 2003 年 10 月 22 日

[11] 授权公告号 CN 1125432C

[22] 申请日 1998.7.22 [21] 申请号 98808942.4

[30] 优先权

[32] 1998. 1. 8 [33] US [31] 09/002,616

[86] 国际申请 PCT/IL98/00341 1998.7.22

[87] 国际公布 WO99/35639 英 1999.7.15

[85] 进入国家阶段日期 2000.3.8

[71] 专利权人 先进识别技术有限公司

地址 以色列特拉维夫

[72] 发明人 耶胡达·赫什科维茨

加布里埃尔·伊兰

审查员 刘红梅

[74] 专利代理机构 永新专利商标代理有限公司

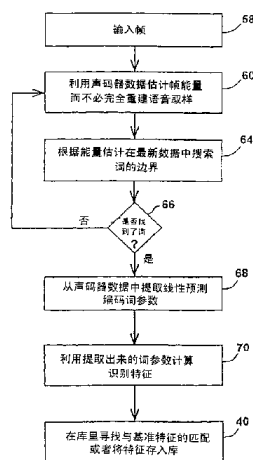
代理人 韩 宏

权利要求书 4 页 说明书 13 页 附图 9 页

[54] 发明名称 基于声码器的语音识别器、识别方法及数字蜂窝电话

[57] 摘要

一种利用基于线性预测编码的声码器数据识别说出来的词而不需要完全重建语音数据的基于声码器的语音识别器。该识别器对于所述声码器数据的每一帧至少生成一个能量估计(60)并利用相关的能量估计搜索所述声码器数据中词的边界(64)。如果找到一个词(66),就从与这一个词相关的声码器数据中提取线性预测编码词参数(68)并从提取出来的线性预测编码词参数计算识别特征(70)。最后,将所述识别特征与以前存储的其它词的识别特征进行匹配(40),从而识别说出来的词。



ISSN 1008-4274

1、一种利用基于线性预测编码（LPC）的声码器数据识别说出来的词而不需要完全重建语音数据的方法，其中的声码器数据构成一系列的帧，该方法包括以下步骤：

对于所述声码器数据的每一帧生成一个或多个能量估计；

利用相关的能量估计搜索所述声码器数据中词的边界；

如果找到一个词，就从与这一个词相关的声码器数据中提取 LPC 词参数；

从提取出来的 LPC 词参数计算识别特征；和

将所述识别特征与以前存储的其它词的识别特征进行匹配，从而识别说出来的词。

2、一种利用基于线性预测编码（LPC）的声码器数据准备识别说出来的词而不需要完全重建语音数据的方法，其中的声码器数据构成一系列的数据帧，该方法包括以下步骤：

对于所述声码器数据的每一帧至少生成一个能量估计；

利用相关的能量估计搜索所述声码器数据中词的边界；

如果找到一个词，就从与这一个词相关的声码器数据中提取 LPC 词参数；

从提取出来的 LPC 词参数计算识别特征。

3、根据权利要求 2 的一种方法，其中所述的生成步骤包括从在所述声码器数据中找到的残差数据估计能量的步骤。

4、根据权利要求 3 的方法，其中估计能量的步骤包括从所述声码器数据重建残差数据并计算所述残差数据的范数的步骤。

5、权利要求3的方法，其中估计能量的步骤包括从所述声码器数据中提取音调-增益值并将提取出来的音调-增益值用作所述能量估计的步骤。

6、根据权利要求3的方法，其中所述生成步骤包括以下步骤：

从所述声码器数据中提取音调-增益值、滞后值和剩余数据；

从所述剩余数据重建剩余信号；

生成所述剩余信号的能量估计；

利用所述音调-增益值和前面用所述滞后值定义的能量估计，生成所述残差的非剩余部分能量估计；和

将所述剩余和非剩余能量估计合并在一起。

7、根据权利要求1的方法，其中声码器数据的类型是以下声码器中任意一个产生的数据的类型：规则脉冲激励-长期预测（RPE-LTP）全速率和半速率、夸尔考姆码激励线性预测（QCELP）8和13 kbps、增强可变速率编解码器（EVRC）、低延迟码激励线性预测（LD CELP）、矢量和激励线性预测（VSELP）、共轭结构代数码激励线性预测（CS ACELP）、增强型全速率声码器和线性预测编码（LPC）10。

8、根据权利要求2的方法，其中所述声码器数据的类型是以下声码器中任意一个产生的数据的类型：规则脉冲激励-长期预测全速率和半速率、夸尔考姆码激励线性预测8和13 kbps、增强可变速率编解码器、低延迟码激励线性预测、矢量和激励线性预测、共轭结构代数码激励线性预测、增强型全速率声码器和线性预测编码10。

9、一种数字蜂窝电话，包括：

一个移动电话操作系统；

一个声码器，利用基于线性预测编码（LPC）的方法压缩语音信号由此产生声码器数据；和

一个基于声码器的语音识别器，包括：

一个前端处理器，用来处理所述声码器数据，以确定一个词是什么时候说出来的，并产生所述说出来的词的识别特征；和

一种识别器，能够将说出来的词识别为一组基准词中的一个。

10、根据权利要求 9 的数字蜂窝电话，其中的前端处理器包括：

一个能量估计器，利用构成所述声码器数据一部分的残差信息估计语音信号的能量；

一个 LPC 参数提取器，用于提取所述声码器数据的 LPC 参数；和

一个识别特征生成器，从所述 LPC 参数生成所述识别特征。

11、根据权利要求 9 的蜂窝电话，其中的前端处理器有选择地与多种声码器一起工作。

12、根据权利要求 9 的蜂窝电话，其中的声码器是以下声码器中的任意一种：规则脉冲激励—长期预测全速率和半速率、夸尔考姆码激励线性预测 8 和 13 kbps、增强可变速率编解码器、低延迟码激励线性预测、矢量和激励线性预测、共轭结构代数码激励线性预测、增强型全速率声码器和线性预测编码 10。

13、一种基于声码器的语音识别器，可以利用基于 LPC 的声码器产生的数据进行工作，该语音识别器包括：

一个前端处理器，用来处理所述声码器数据，以确定一个词是什么时候说出来的，并产生所述说出来的词的识别特征；和

一种识别器，至少能够将所说出来的词识别为一组基准词中的一个。

14、根据权利要求 13 的语音识别器，其中的前端处理器包括：

一个能量估计器，利用构成所述声码器数据一部分的残差信息估计语音信号的能量；

一个 LPC 参数提取器，用于提取所述声码器数据的 LPC 参数；和
一个识别特征生成器，从所述 LPC 参数生成所述识别特征。

15、根据权利要求 14 的语音识别器，其中的能量估计器包括一个残差能量估计器，用来从在所述声码器数据中找到的残差数据估计能量。

16、根据权利要求 15 的语音识别器，其中的残差能量估计器包括一个残差数据重建器，用来从所述声码器数据重建残差数据，还包括一个范数生成器，它能生成所述残差数据的范数并产生所述能量估计。

17、根据权利要求 15 的语音识别器，其中的残差能量估计器包括一个提取器，它能从所述声码器数据中提取音调-增益值，并产生所述能量估计。

18、根据权利要求 15 的语音识别器，其中的残差能量估计器包括：
从所述声码器数据中提取音调-增益值、滞后值和剩余数据的一个提取器；

从所述剩余数据重建剩余信号的一个重建器；

产生所述剩余信号的能量估计的一个剩余能量估计器；

一个非剩余能量估计器，利用所述音调-增益值和所述滞后值定义的前面的能量估计，产生所述残差数据的非剩余部分的一个能量估计；
和

合并所述剩余能量估计和非剩余能量估计从而产生所述能量估计的一个合并器。

19、根据权利要求 13 的语音识别器，其中的声码器是以下声码器中的任意一种：规则脉冲激励-长期预测全速率和半速率、夸尔考姆码激励线性预测 8 和 13 kbps、增强可变速率编解码器、低延迟码激励线性预测、矢量和激励线性预测、共轭结构代数码激励线性预测、增强型全速率声码器和线性预测编码 10。

基于声码器的语音识别器、识别方法及数字蜂窝电话

发明领域

本发明总地涉及语音识别器，并涉及使用线性预测编码（LPC）声码器数据作为输入的语音识别器。

发明背景

在本领域里，语音识别器大家都非常熟悉，它在许多地方都有应用。例如，移动装置的指令和控制应用、计算机录音电话机、儿童玩具和汽车电话里都使用了语音识别。在所有这些系统里，都将语音信号数字化，然后参数化。将参数化的输入信号与其发音已知的基准参数化信号比较。与输入信号最吻合的基准信号的相关发音就是识别出来的发音。

语音识别系统在语音拨号系统中有其特殊的用途，当用户说出他想呼叫的人的名字时，语音识别系统就从事先提供的基准表识别出这一名字，并给出这一识别出来的名字所对应的电话号码。然后电话机拨这一号码。结果是用户接通了他想要的电话，而不必寻找电话号码，也不必用自己的手来拨电话号码。

语音拨号对汽车移动电话来说特别重要，在这种情况下，电话用户通常就是汽车驾驶员，他必须不间断地将注意力集中在路面上。如果驾驶员想给某个人打电话，驾驶员说出这个人的名字比他亲自去拨电话号码要安全得多。

下面参考图 1, 它说明的是一架数字移动电话的主要部件。一般而言, 移动电话中有一个麦克风 10、一个扬声器 12、一个模数转换单元 14、一个用标为 DSP-1 的数字信号处理 (DSP) 芯片实现的声码器 16、微控制器或中央处理单元 (CPU) 中的一个操作系统 18、一个射频接口单元 (RF 单元) 19 和一付天线 20。发射信号的时候, 麦克风 10 产生模拟语音信号, 由单元 14 数字化。声码器 16 压缩语音取样, 以减少要通过射频单元 19 和天线 20 发射给另一个移动电话的数据量。接收端移动电话的天线 20 通过射频单元 19 将收到的信号提供给声码器 16, 声码器 16 对收到的信号解压缩, 得到语音取样。单元 14 将语音取样转换成模拟信号, 由扬声器 12 播放出来。操作系统 18 控制移动电话的工作过程。

对于语音拨号系统, 移动电话还有一个语音识别器 22, 实现在标为 DSP-2 的一个单独的 DSP 芯片里, 它接收数字化的语音取样作为输入, 对该语音信号进行参数化, 并将参数化的输入信号与基准语音信号比较。一般情况下, 语音识别器 22 或是将匹配信号的识别码提供给操作系统 18, 或者是提供有关的电话号码, 如果这一匹配信号有一个电话号码与之对应。

现在参考图 2, 它说明语音识别器 22 工作的总过程。将数字化的语音取样按预定长度分成帧, 例如长度为 5—20 ms 的帧, 提供给识别器 22 的就是这些帧 (步骤 28)。对于每一帧, 识别器 22 首先计算 (步骤 30) 这一帧的能量。

现在参考图 3, 它说明的是, 作为时间的函数, 说出来的词“RICHARD”的每一帧的能量。这一能量信号有两个峰 31 和 33, 对应于这个词的两个音节。不说话的时候, 图中用参考数字 35 表示, 甚至在音节之间, 能量电平显然要低得多。

因此, 识别器 22 在能量信号里搜索 (图 2 中的步骤 32) 一个词的

开头和结尾。将词的开头定义为能量信号维持低电平的时间超过一段预定长度以后，能量开始显著增大的那一个点 37。将词的结尾定义为能量的显著降低结束的那一点 39，在这一点以后，能量信号维持低电平的时间比一段预定时间的长度长。在图 3 里，词的开头 37 大约在 0.37 s 处，结尾 39 大约在 0.85 s 处。

如果找到了一个词，见步骤 34 中选中的分支，语音识别器 22 就进行（步骤 36）线性预测编码（LPC）分析，从而生成所说词的参数。在步骤 38 里，语音识别器 22 计算说出的词的识别特征，在步骤 40 里，语音识别器 22 从基准库里搜索匹配的基准词识别特征。或者语音识别器 22 在叫做“训练”的过程里将识别特征存入基准库。

可惜，语音识别过程计算量很大，因此只能在第二 DSP 芯片 DSP-2 里才能实现。这给移动电话增加了不少的成本。

发明概述

本发明的目的是提供一种语音识别器，它处理的是由基于 LPC 的声码器压缩过的语音数据，而不是语音数据取样，从而减少识别器的计算量。这样，语音识别可以在同时运行操作系统程序的微控制器或 CPU 中进行。由于语音识别器不分析语音信号，就可以用处理能力有限和/或不需要接收语音信号的微控制器或者 CPU。

此外，本发明提供一种特征发生器，它可以从基于 LPC 的不同类型的声码器提取相同类型的特征数据用于识别。这样，本发明可以针对用基于 LPC 的不同类型的声码器压缩过的压缩语音数据，进行同样的识别操作（例如比较和训练）。

因此本发明的一个优选实施方案提供了一种方法，用于利用基于线性预测编码（LPC）的声码器数据识别说出来的词，而不必完全重建语

音数据。这一基于声码器的识别器用于实现这里描述的方法。该方法包括对声码器数据产生至少一个每帧能量估计，以及利用相关的能量估计从声码器数据中搜索词的边界这两个步骤。如果找到了一个词，就从与这一个词相关的声码器数据中提取出 LPC 词参数，并从提取出来的 LPC 词参数中计算出识别特征。最后将识别特征与先前存储的其它词的识别特征比较，从而识别说出来的词。

此外，根据本发明的一个优选实施方案，能量是从在码器数据中找到的残差数据估计出来的。可以用很多的方法来进行这一估计。在一个实施方案中，从声码器数据中重建出残差数据，得到的估计就是残差数据的范数。另一个实施方案是从声码器数据中提取一个音调-增益 (pitch-gain) 值，作为能量估计。在再一个实施方案里，从声码器数据中提取出音调-增益值、滞后值 (lag values) 和剩余数据 (remnant data)。从剩余数据中产生剩余信号，由此产生一个剩余能量估计。非剩余能量估计是利用音调-增益值和滞后值所定义的前一个能量估计值从残差信号的非剩余部分产生的。最后将这两个能量估计，剩余能量和非剩余能量合并。

更进一步，根据本发明的一个优选实施方案，声码器数据可以来自以下声码器中的任何一个：RPE-LTP 全速率或半速率，QCELP 8 和 13 kbps，EVRC，LD CELP，VSELP，CS ACELP，增强型全速率声码器和 LPC10。

本发明的又一个优选实施方案提供了一种数字蜂窝电话，它包括一个移动电话操作系统、一个基于 LPC 的声码器和一个基于声码器的语音识别器。该识别器包括一个前端处理器，对声码器数据进行处理以确定一个词是什么时候说出来的，并产生这一说出来的词的识别特征，该识别器还包括一个识别器，它至少能够识别出说出来的词是一组基准词中的哪一个。

根据本发明的另一个实施方案，前端处理器有一个能量估计器、一个 LPC 参数提取器和一个识别特征发生器。能量估计器利用构成声码器数据一部分的残差信息估计语音信号的能量。LPC 参数提取器提取声码器数据的 LPC 参数。识别特征生成器从 LPC 参数生成识别特征。

更进一步，根据本发明的一个优选实施方案，前端处理器可以有选择性地与多种声码器一起工作。

附图简述

通过下面的详细描述并参考以下附图，可以获得对本发明更全面的理解。

图 1 是现有技术中有语音识别功能的蜂窝电话的一个框图；

图 2 是现有技术中基于 LPC 的语音识别方法的流程图；

图 3 是一个说出来的词的能量图；

图 4 是压缩语音数据的结构示意图；

图 5 是按照本发明的一个优选实施方案构成和工作的有一个基于声码器的语音识别器的一个蜂窝电话的框图；

图 6 是本发明一个优选实施方案中一种语音识别方法的流程图；

图 7 是从残差信号中估计出来的一个说出来的词的能量图；

图 8 是有利于理解本发明工作过程的残差信号的示意图；

图 9 是 GSM 解码器的一个框图；和

图 10 是从估计的残差信号中估计出来的一个说出来的词的能量图。

发明详述

有很多种语音压缩算法，最常用的是那些基于线性预测编码（LPC）的算法。本发明的申请人认识到，由于多数语音识别算法都是利用线性

预测编码分析以便从语音信号中提取参数，就可以将压缩语音信号的元素提供给语音识别器，以便显著地降低语音识别器的分析量。因此，本发明是一个实现在蜂窝移动电话的微控制器或 CPU 中的基于声码器的语音识别器，下面将参考图 5、6 和 7 详细说明。

线性预测分析：

下面简要说明基于 LPC 的声码器 16 的工作原理。对语音编码的一般性讨论，与本说明相比，包括对线性预测编码更加详细的说明，可以在 Andreas S. Spanias 在电气与电子工程师学会会报，1994 年 10 月，第 82 卷第 10 期，第 1541—1582 页上发表的文章“语音编码：综述”中找到。

声码器 16 将语音信号分成一系列的帧，每一帧的长度为 N，通常包括大约 20 ms 的语音信号。声码器 16 对每一帧进行线性预测编码（LPC）分析。

线性预测编码用以下方程描述语音信号 $y(n)$ ：

等式 1

$$y(n) = a_1 y(n-1) + a_2 y(n-2) + \dots + a_p y(n-p) + \varepsilon(n)$$

其中 a_i 叫做 LPC 系数， $\varepsilon(n)$ 叫做残差信号。每一帧通常都有 p 个 LPC 系数 a_i ，残差信号 $\varepsilon(n)$ 的长度是 N。LPC 系数和残差信号构成帧的参数。声码器通常还至少在音调和增益值方面对残差信号 $\varepsilon(n)$ 进行参数化。声码器还可以产生 LPC 声码器领域中众所周知的基于 LPC 的许多参数中的任意一种，例如倒频谱系数、MEL 倒频谱系数、线谱对（LSP）、反射系数、对数面积比（LAR）系数等等，所有这些都可以很容易地从 LPC 系数计算出来。

接着对获得的值进行编码，从而产生典型的语音压缩帧，例如图 4

所示的帧 52。语音压缩帧 52 包括编过码和/或参数化的 LPC 系数 a_i 和编过码的残差信号 $\varepsilon(n)$ 。

基于声码器的语音识别器

现在参考图 5，它说明的是蜂窝电话中一个基于声码器的语音识别器 50。由于蜂窝电话与图 1 所示现有技术电话相似，因此图中相同的数字就表示相同的单元。还要参考图 6 和 7，它们有助于理解基于声码器的语音识别器 50 的操作。

图 5 中的蜂窝电话包括麦克风 10、扬声器 12、转换单元 14、声码器 16、操作系统 18、射频接口单元 19 和天线 20。另外，图 5 中的蜂窝电话包括基于声码器的语音识别器 50，用于接收声码器 16 产生的基于 LPC 的压缩语音信号作为输入。

根据本发明的一个优选实施方案，基于声码器的语音识别器 50 是在 CPU 51 里实现的，CPU 51 还实现操作系统 18。装置 51 可以是一个 CPU，就象所标明的那样，或者是一个微控制器。由于语音识别器 50 不分析语音信号，就可以用任意一种微控制器或者 CPU 来实现语音识别器 50，包括处理能力有限并且不接收语音信号的那些微控制器或者 CPU。

图 6 说明基于声码器的语音识别器 50 处理象帧 52 那样的压缩帧的基本形式。

象在现有技术中那样，一旦收到帧（步骤 58）就确定帧的能量。而在本发明中，能量不是从取样数据而是从声码器数据中估计出来的（步骤 60），而且能量估计不需要重建取样数据。

本发明的申请人发现可以利用残差信号 $\varepsilon(n)$ 来估计能量，因为在本领域里大家都知道，残差信号描述的是通过声道的空气的压力，而 LPC 参数描述的则是声道的结构因而通常与语音音量无关。结果是，残差信号与一个人说话的声音是大还是小密切相关。

根据本发明的一个优选实施方案，估计能量的一种方法是确定每一帧残差信号的能量，或者如果帧分成了子帧，就是确定每一子帧的能量。这用数学公式表示就是：

等式 2

$$\tilde{E}_i = \sum_{n=1}^M \varepsilon(n)^2$$

其中 $\tilde{E}_i$ 是第 i 帧的能量，残差信号 $\varepsilon(n)$ 是从声码器数据中重建出来的，数字 M 是帧或子帧中的取样点数。

图 7 说明的是从说出来的词“RICHARD”的重建残差信号得到的估计的能量信号。可以看出，图 7 中的估计出来的能量信号与图 3 中的能量信号并不完全相同。但估计出来的能量信号与现有技术中的能量信号相关性非常好。图 7 中信号的开头和结尾，分别标成 62 和 63，也都是在大约 0.37 s 和 0.85 s 处。

本发明还包括从声码器数据估计能量的其它方法，下面将介绍其中的一些。

回到图 6，基于声码器的语音识别器 50 在估计的能量信号中搜索（步骤 64）词的边界。如果需要，语音识别器 50 可以利用 LPC 参数的任何特征（例如它们的平均值和/或方差）来确定词的边界位置，这些参数的特征在词的边界上急剧地变化。

如果找到一个词，就象步骤 66 中检验出来的一样，识别器 50 就从声码器数据中提取出 LPC 词参数（步骤 68）。在步骤 68 里通常还要对语音压缩帧 52 中的编码 LPC 参数解码并将它们转换成 LPC 系数。

然后识别器 50 从提取出来的 LPC 系数中计算（步骤 70）它的识别特征。这些识别特征可以是基于 LPC 的参数中的任意一个，例如倒频谱系数、MEL 倒频谱系数、线谱对（LSP）、反射系数、面积比（LAR）系数等等，所有这些都很容易地从 LPC 系数计算出来。这样，如果

声码器使用的是一种 LPC 参数, 识别器 50 使用的是另一种 LPC 参数, 识别器 50 就可以直接或者通过 LPC 系数将 LPC 参数从一种转换成另一种。

最后识别器 50 利用从步骤 70 获得的识别特征将输入信号识别成它的基准库里的一个基准词, 或者训练出一个新的基准词存入它的库中。由于识别器 50 产生的识别特征可以与本领域里的一样, 这一步骤与现有技术里的识别/训练步骤 40 一样, 它也是这样标注的。Prentice-Hall 1993 年出版的 Lawrence Rabiner 和 Bing Hwang Juang 的书《语音识别基础》描述了许多合适的识别器 50, 这里将它引做参考。

应当理解, 步骤 60—70 是将声码器数据转换成识别/训练步骤所需要的识别特征。有许多种基于 LPC 的声码器, 每一种对语音信号的操作多少都有些不同。可以将步骤 60—70 修改以适用于每一种声码器, 以便产生相同的识别特征, 而不考虑声码器类型。因此步骤 60—70 构成识别/训练步骤 40 的处理“前端”。

本发明包括一种基于声码器的语音识别器, 它有多个前端和一个识别/训练单元。这对于卖出去用于在多种数字蜂窝电话系统中工作的那些移动电话非常有用, 这里的每一种蜂窝电话系统都使用不同种类的声码器。有了许多前端, 本发明的语音识别器就可以与许多种声码器协同工作。

用于确定词的边界的能量估计方法:

一些简单的声码器, 例如美国国防部标准 1015 V. 53 里描述的叫做线性 LPC 10 的声码器, 只用音调和增益值描述残差信号 $\varepsilon(n)$ 。现在参考图 8, 其中是一个浊音信号的残差信号实例, 该信号有许多重复的尖峰 70, 这些尖峰 70 的幅度都差不多。将相邻两个尖峰 70 之间的距离定义为音调 (pitch) P, 这些尖峰的幅度定义为增益 G。清音信号有增益值,

但没有音调值。

因此，帧或子帧残差信号能量可以用增益值 G 来估计。在这一个实例里，不是通过重建残差信号 $\varepsilon(n)$ ，而是通过从压缩语音数据中提取残差信号 $\varepsilon(n)$ 的一个参数：增益值 G ，来估计帧或子帧的能量。

其它的声码器，例如全球移动通信系统（GSM）、时分多址（TDMA）和码分多址（CDMA）数字蜂窝通信系统中的声码器，将当前帧或子帧的残差信号与前面一些帧的残差信号的并置版本做相关运算。当前帧的残差信号与前面的残差信号最相近的点，与音调增益（pitch gain） PG 相乘以后，叫做 LAG 值。然后声码器确定一个“剩余信号”，它是前一个残差信号乘以音调增益 PG 与当前残差信号的差。于是当前残差信号的特征是音调增益 PG 、 LAG 值和剩余信号。

对于后一种声码器，当前帧或子帧， i ，的能量可以利用以下公式从剩余信号和残差信号的非剩余部分估计出来：

等式 3

$$\tilde{E}_i = \sqrt[m]{E_{LAG}^m + E_{rem}^m} \quad (m=1 \text{ 或者 } 2)$$

$$E_{LAG} = PG \cdot \frac{1}{FL} \left\{ (LAG \bmod 40) E_{i-\left\lceil \frac{LAG}{FL} \right\rceil} + (FL - LAG \bmod 40) E_{i-\left\lfloor \frac{LAG}{FL} \right\rfloor} \right\}$$

其中 E_{rem} 是剩余信号的能量估计， E_{LAG} 是残差信号的非剩余能量，它是从比当前帧或子帧晚 $\frac{LAG}{FL}$ 帧或子帧的帧或子帧的能量和音调增益获得的。可以通过重建剩余信号来获得前一个值，这一运算相对简单，或者通过任何其它方法来获得前一个值。符号 $\lceil \cdot \rceil$ 和 $\lfloor \cdot \rfloor$ 分别表示“不大于变量的最大整数”和“不小于变量的最小整数”运算，开 m 次方运算不必进行。

GSM 声码器的能量估计

现在参考图 9，其中说明的是构成部分 GSM 标准的声码器的解码器部分。图 9 与欧洲电信标准协会 1992 年 3 月份版的 I-ETS 300036 规范

中, 第 34 页的图 3、4 类似。解码器的细节可以在该规范中找到, 在这里将该规范引做参考。为清楚起见, 下面只说明解码器中对理解本发明的能量和特征计算有用的有关部分。

图 9 用粗线表示输入数据, 用细线表示内部信号。输入数据有 M_{cr} 、 X_{maxcr} 、 X_{mcr} 、 b_{cr} 、 N_{cr} 和 LAR_{cr} 值, 它们在 I-ETS 规范里都有定义。

图 9 表明这一解码器有一个 RPE 解码器 80、一个长期预测器 84、一个短期合成滤波器 86 和一个去加重器 88。RPE 解码器 80 接收 M_{cr} 、 X_{maxcr} 和 X_{mcr} 信号, 并产生一个剩余信号 e'_r 。长期预测器 84 利用 b_{cr} 和 N_{cr} 信号从剩余信号 e'_r 产生一个残差信号 d'_r 。短期合成滤波器 86 从残差信号 d'_r 和以 LAR_{cr} 数据形式发送的短期 LPC 参数产生语音信号。

与前面一样, 有一种能量计算方法是计算残差信号 d'_r 的第一或第二范数: 如下所示

等式 4

$$\tilde{E}_i = \sum_{n=0}^{39} |d'_r[n]|^m \quad (m=1 \text{ 或者 } 2)$$

另一个能量计算方法是利用剩余信号 e'_r 和长期预测器 84 的内部数据值 b'_r 和 N'_r 。具体而言, 预测器 84 有一个参数解码器 90、一个延迟单元 92、一个乘法器 94 和一个加法器 96。解码器 90 将输入值 b_{cr} 和 N_{cr} 变换成内部数据值 b'_r 和 N'_r , 这里的 b'_r 是一个乘数, 与前面的音调增益 PG 相似, N'_r 是一个延迟量, 与前面的 LAG 值相似。长期预测器 84 将信号 d'_r 加到剩余信号 e'_r 上去, 其中, 信号 d'_r 是前面的残差信号 $d'_r(1-N'_r)$ 经过延迟单元 92 延迟以后, 再在乘法器 94 里用 b'_r 相乘以后得到的结果。

可以用等式 3 估算能量, 其中的 LAG 和 PG 值用 N'_r 和 b'_r 代替, 并将 FL 的值取为 40。另外, 剩余的能量估计 E_{rem} 用以下等式计算:

等式 5

$$E_{rem} = \sum_{k=0}^{39} |e'_r[k]|^m$$

现在参考图 10，它说明的是利用上述公式估计出来的能量。词的开头和结尾两个边界，分别标为 98 和 99，与现有技术里的位置相同。

从提取出来的参数估计能量的另一种方法也是利用前面的 N'_l 和 b'_l ，其中的 FL 取值 40，用以下公式估计剩余的能量估计， E_{rem} ：

等式 6

$$E_{rem} = |X_{max cr}|^m$$

回到图 9，LPC 的词参数是从短期合成滤波器 86 内的发送数据里提取出来的，该滤波器中有一个 LAR 解码器 100、一个内插器 102、一个反射系数确定单元 140 和一个滤波器 106。单元 100、102 和 140 一起将收到的 LAR_{cr} 数据转换成反射系数 r'_l ，再将反射系数容易地变换成 LPC 系数。

就象这里参考图 6 所做的介绍一样，一旦提取出 LPC 系数，就将它们变换（步骤 70）成识别器/训练步骤需要的识别特征。

本领域里的技术人员应当明白，尽管只针对 GSM 数字蜂窝通信系统里的声码器做了详细的介绍，但是本发明仍然可以用于所有类型的蜂窝通信系统和所有类型的基于 LPC 的声码器。对于每一种声码器，必须分析存储在压缩语音数据里的信息类型，以确定计算能量和特征时如何利用它。压缩语音数据的详细介绍可以参见定义每一个声码器的标准。

下表列出了一些现有的蜂窝通信系统、它们使用的声码器和定义声码器和/或系统的标准。

数字蜂窝通信系统	基于 LPC 的声码器	标准
GSM	RPE-LTP 全速率	I-ETS 300 036 6.1
		I-ETS 300 581-2 第 4
	RPE-LTP 半速率	版

CDMA	QCELP 8 kbps, 13 kbps	IS 96 A
	EVRC	IS 127
	LD CELP	ITU G.728
TDMA	VSELP	IS 54 B
PHS, PCS	CS ACELP	ITU G.729
PCS-TDMA	增强型全速率声码器	IS 641
PDC (日本)	VSELP	RCR STD 27

本领域的技术人员应当明白，本发明并不局限于上面的具体描述。
相反，本发明的范围是由以下权利要求规定的。

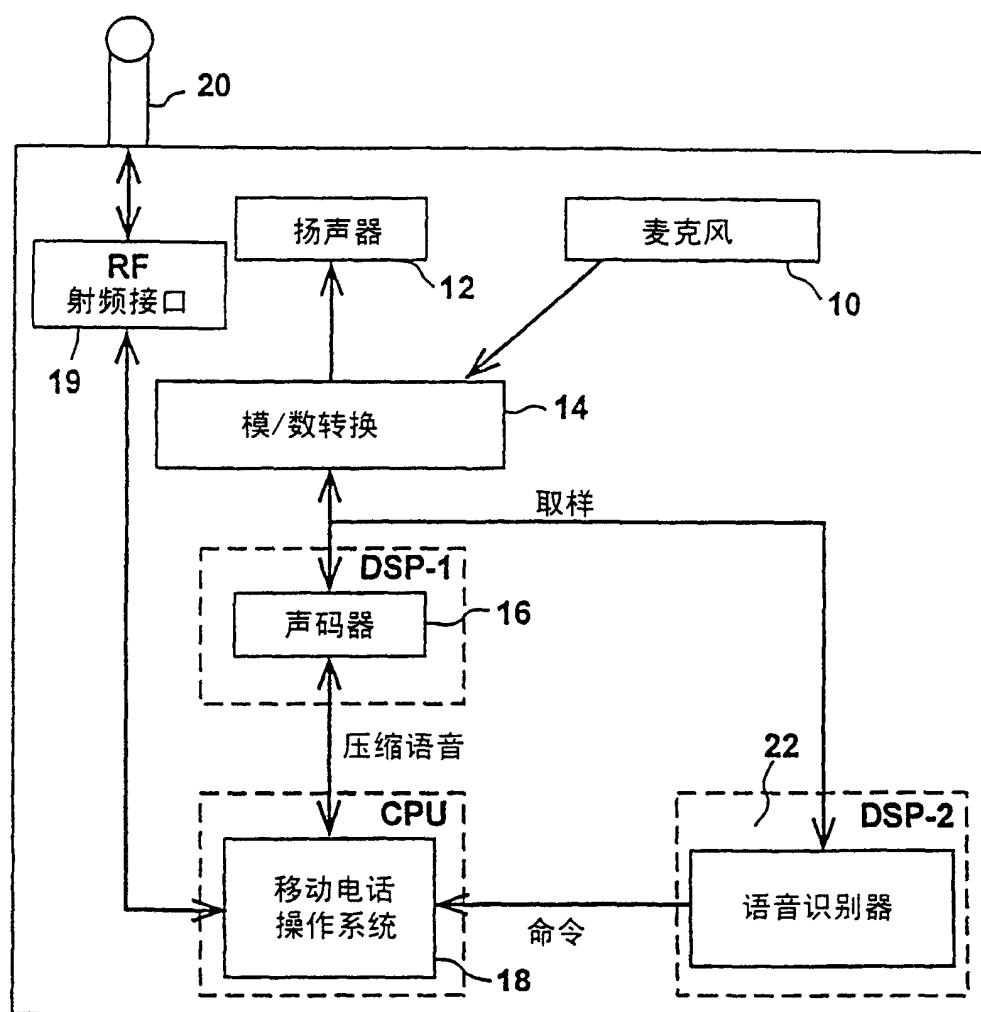


图1

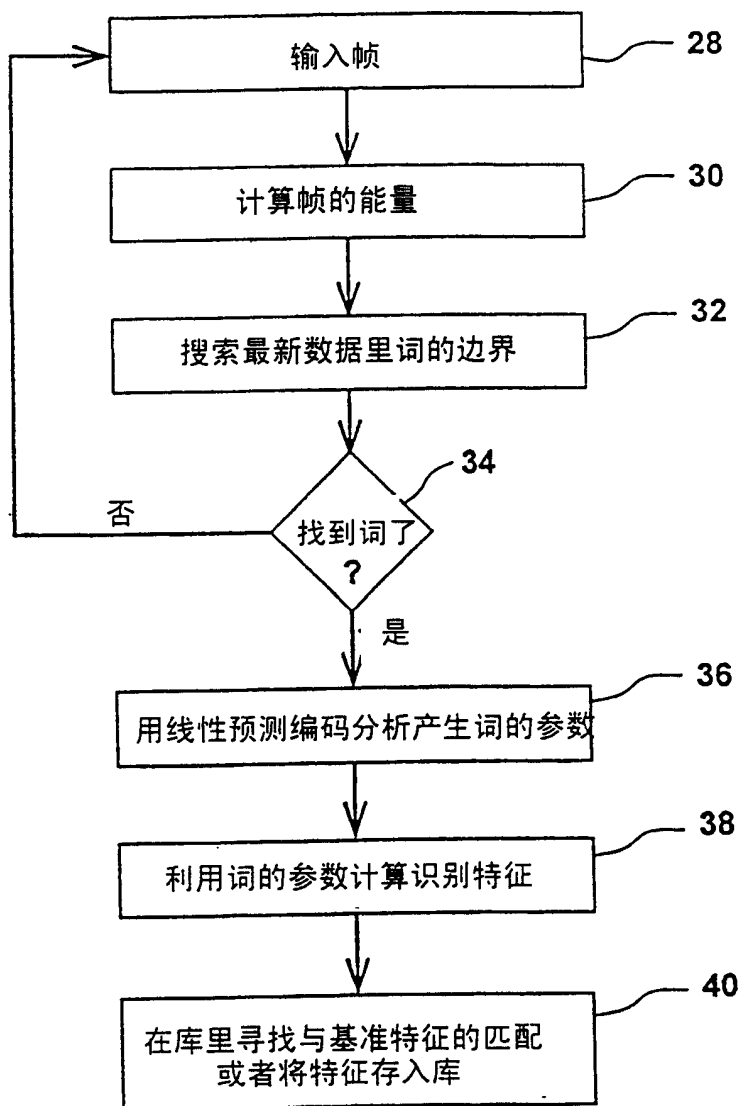


图2

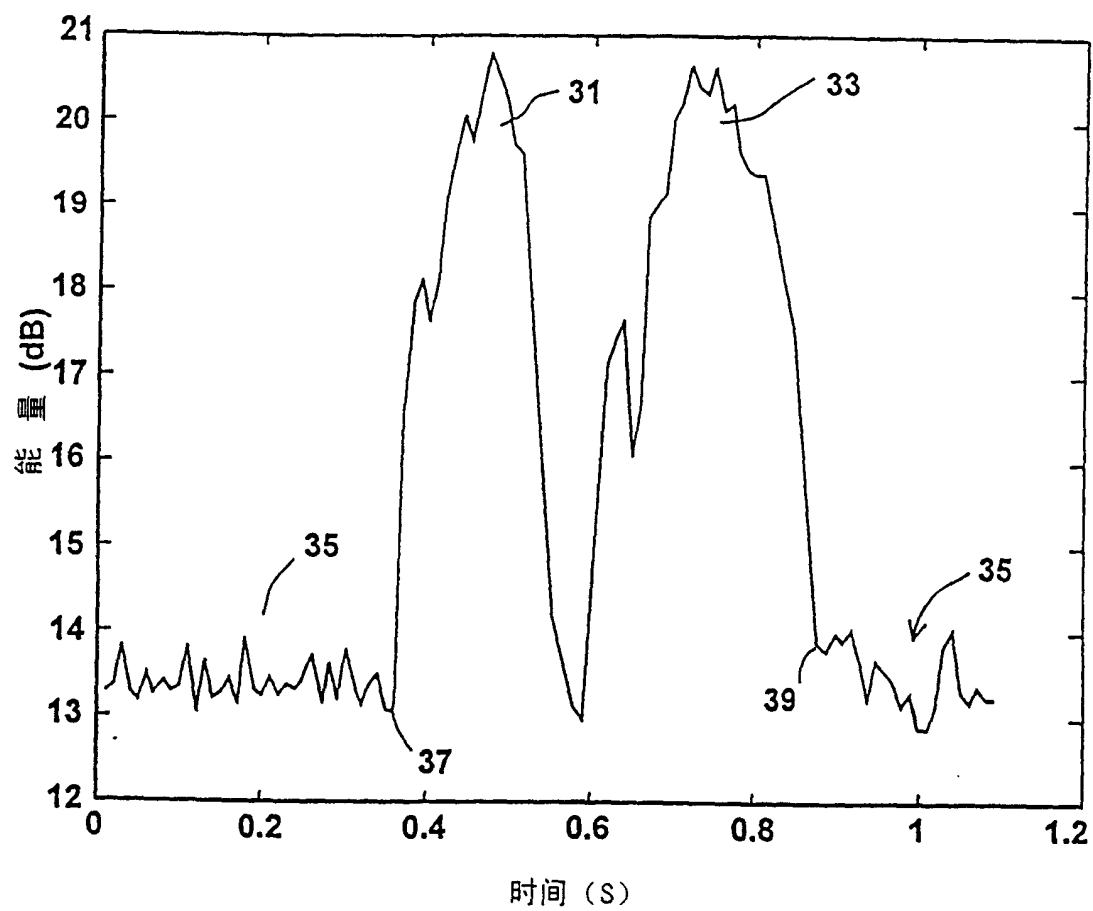


图3



图 4

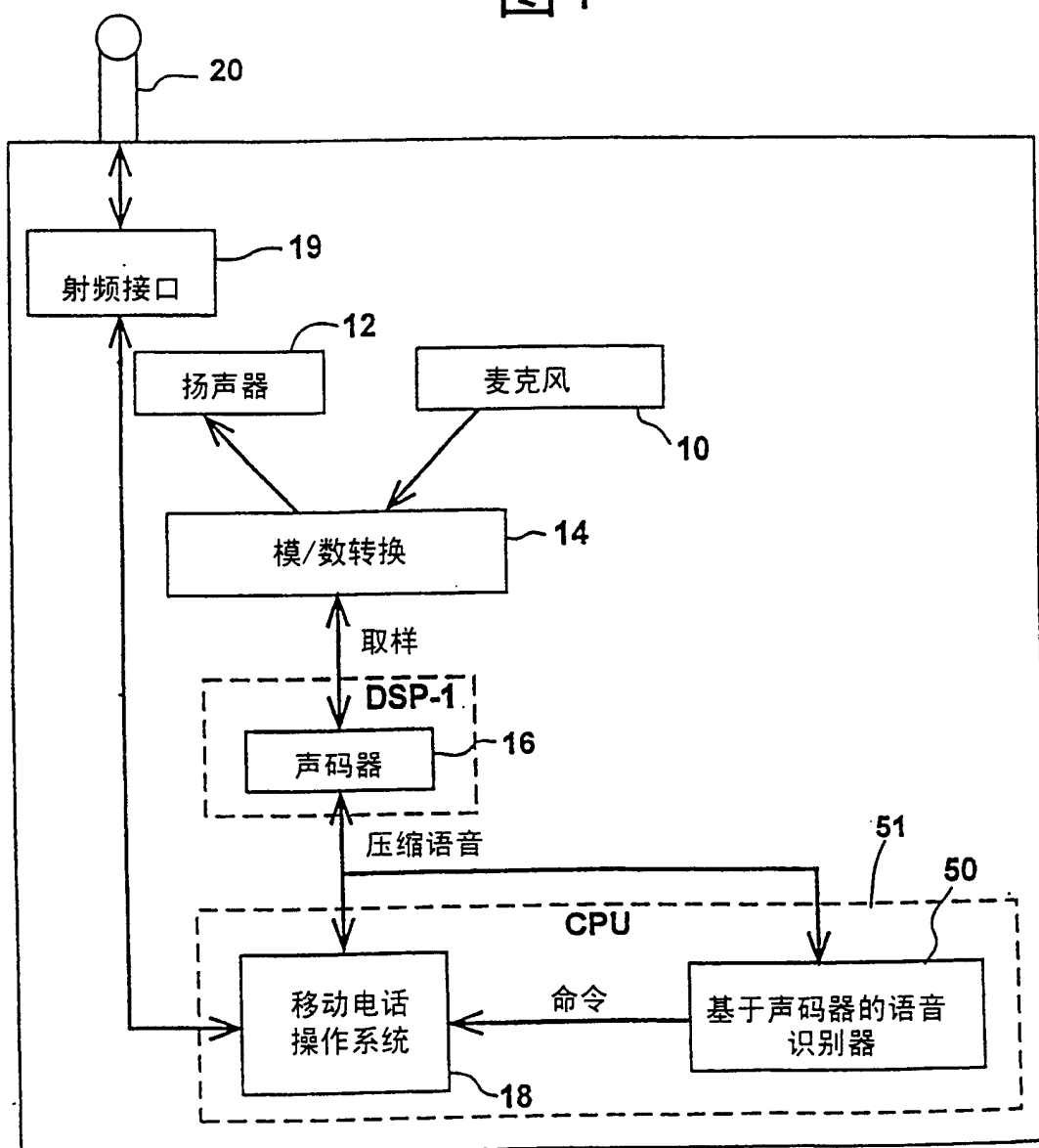


图5

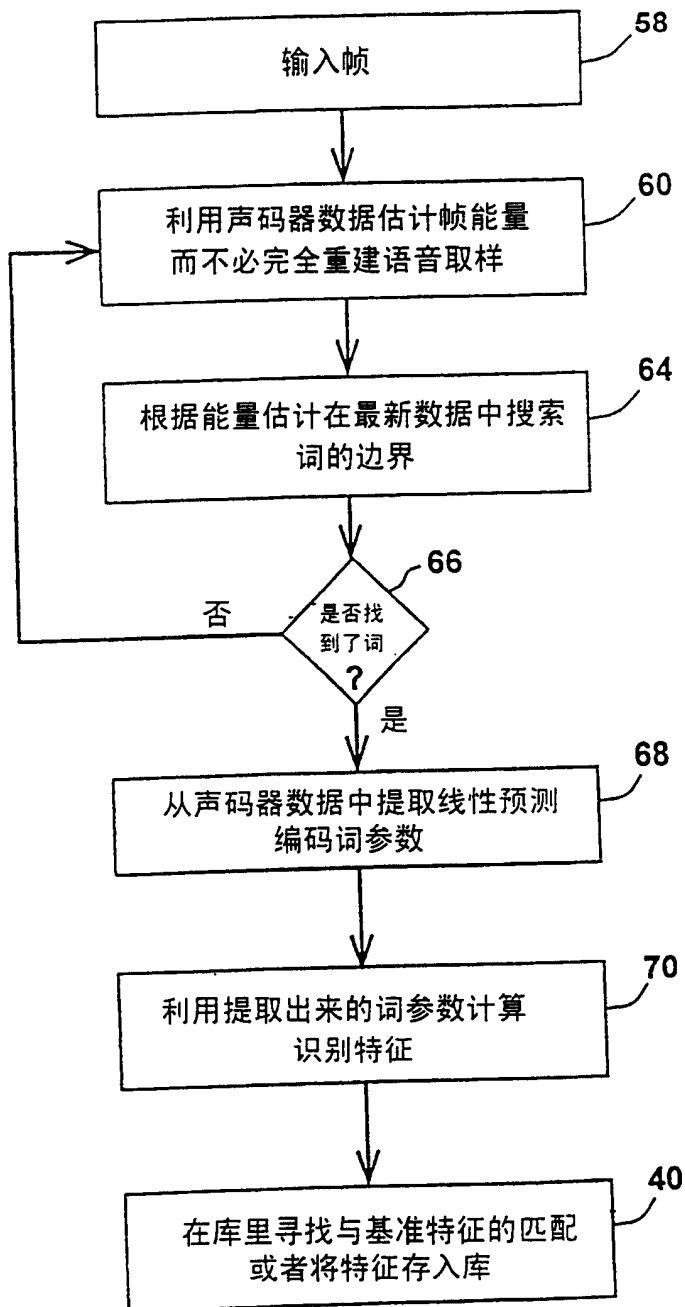


图6

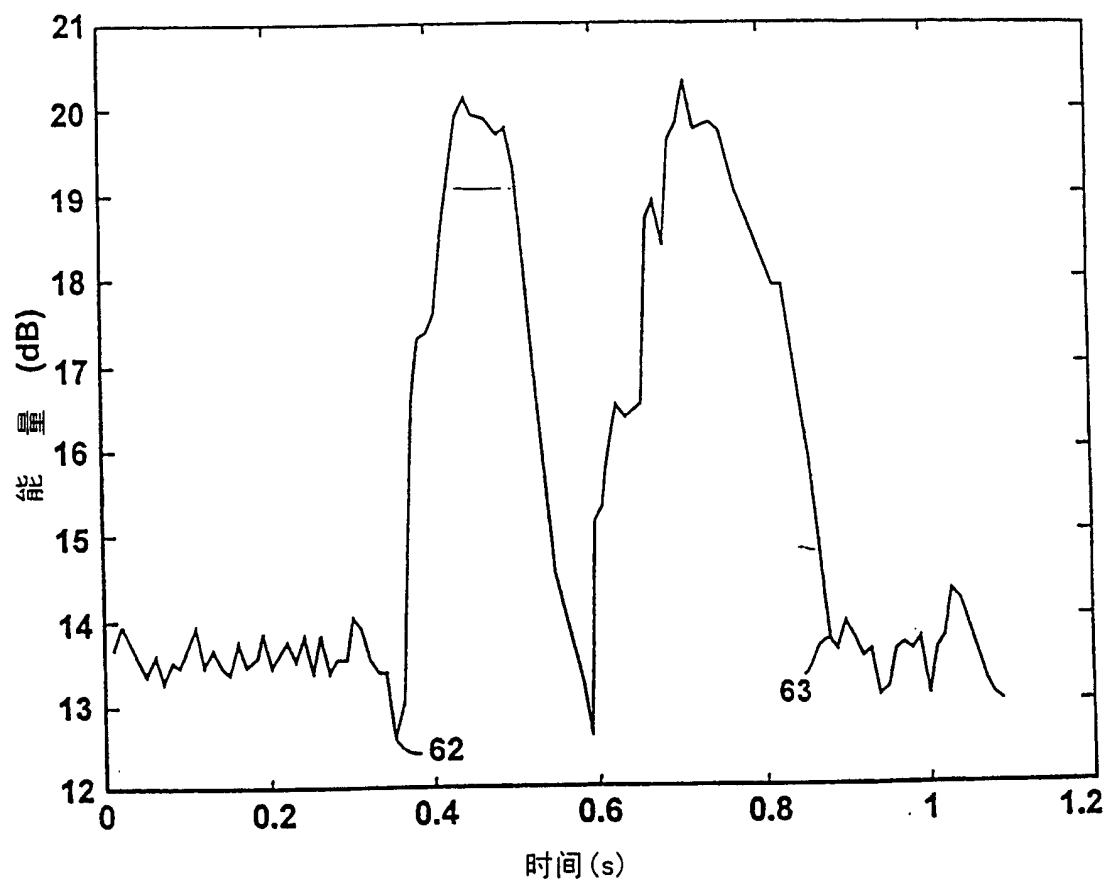


图7

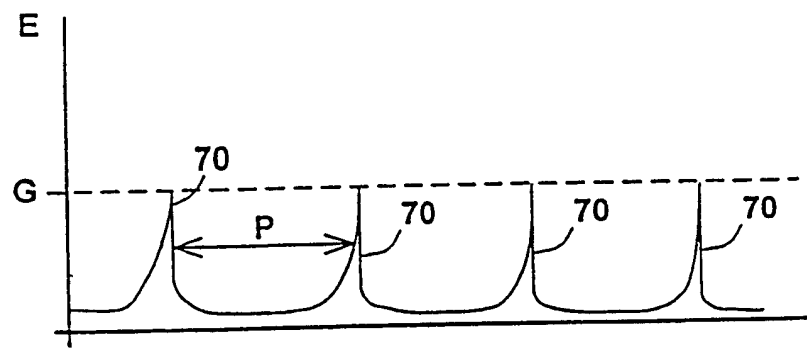


图8

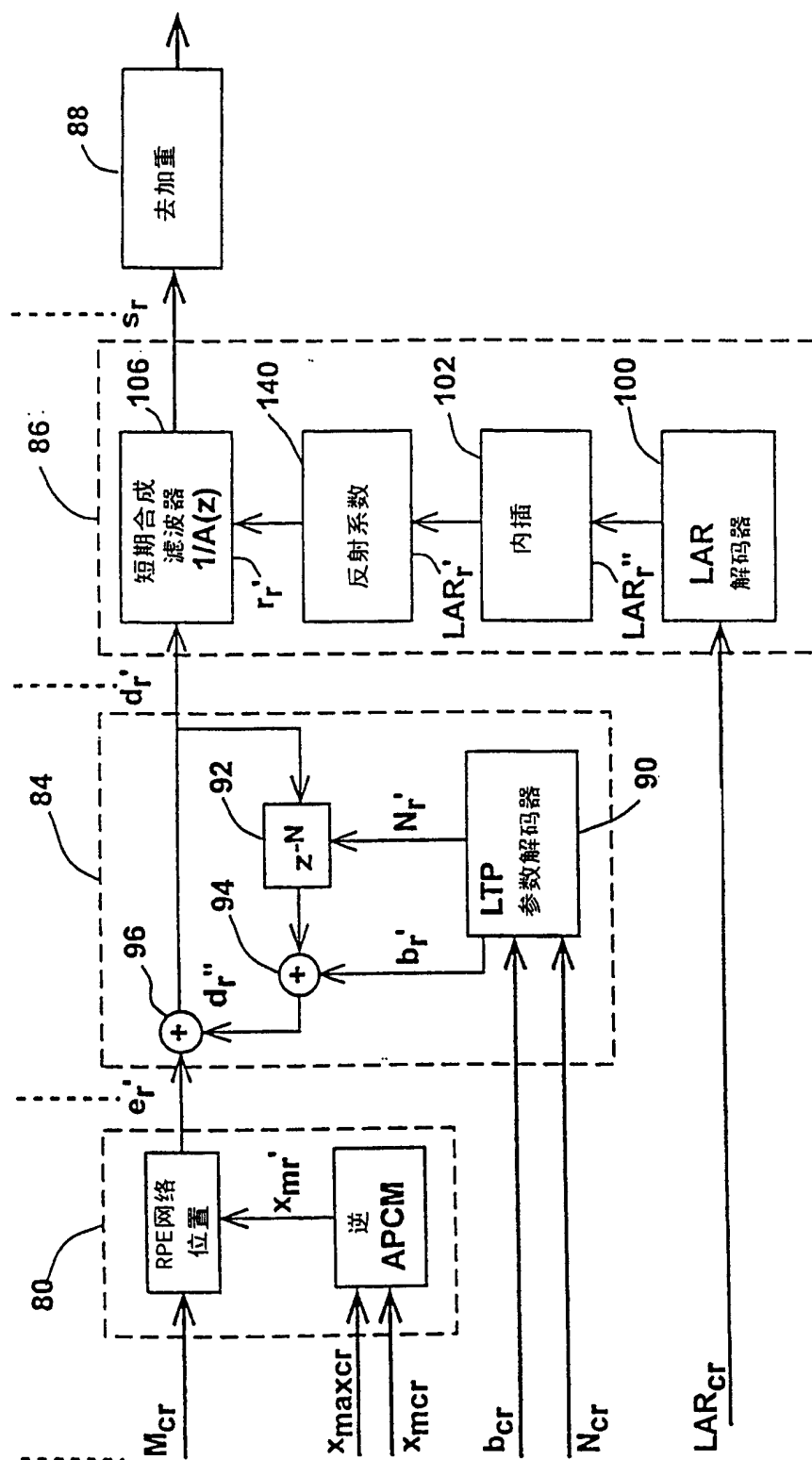


图9

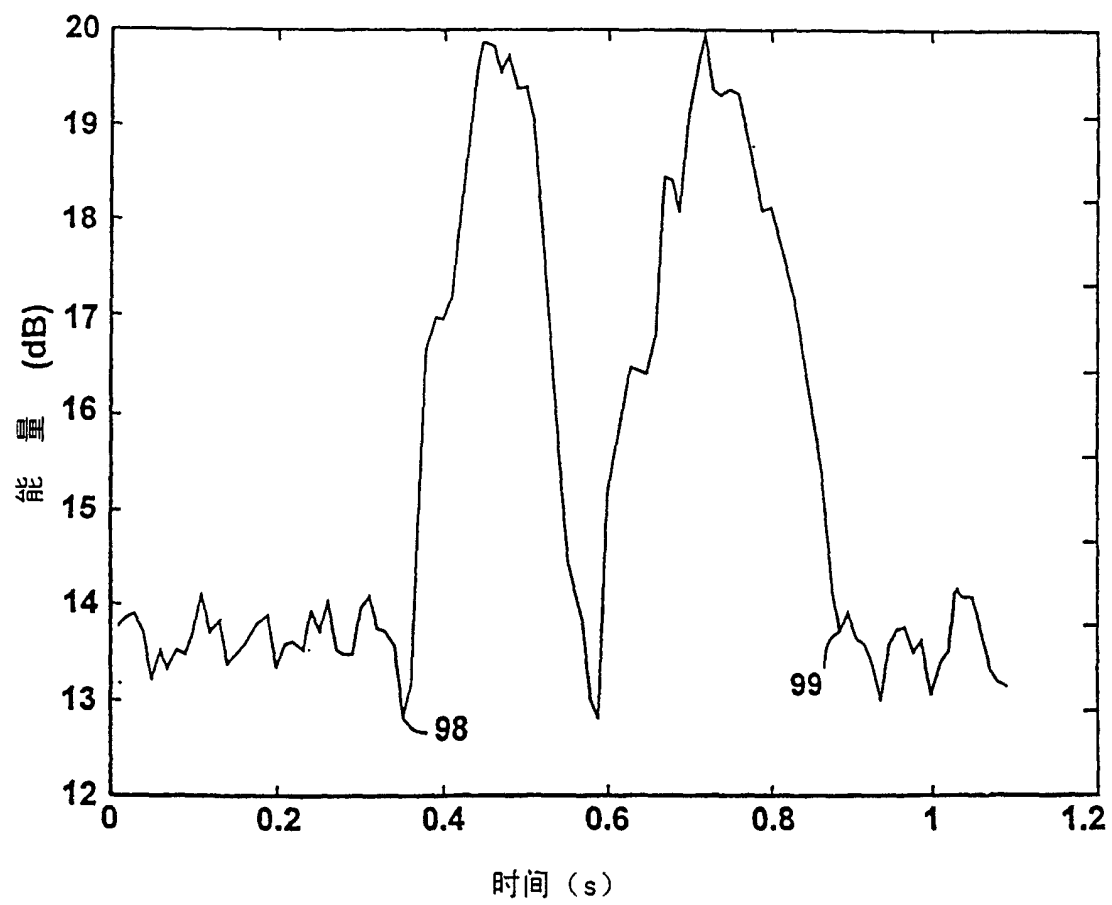


图10