(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 1815558 B
(45)授权公告日 2010.09.29

(21)申请号 200410045610.X
(22)申请日 1999.11.12
(30)优先权数据
09/191,633 1998.11.13 US
(62)分案原申请数据
99815573.X 1999.11.12
(73)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 A·达斯 S·曼朱那什
(74)专利代理机构 上海专利商标事务所有限公司 31100
代理人 李玲
(51)Int.Cl.
G10L 19/08(2006.01)
G10L 19/14(2006.01)
(56)对比文件
CN 1131473 A,1996.09.18, 全文.
US 5517595 A,1996.05.14, 全文.
US 5327521 A,1994.07.05, 全文.
US 5734789 A,1998.03.31, 全文.
US 5490230 A,1996.02.06, 全文.
审查员 邢雲峰

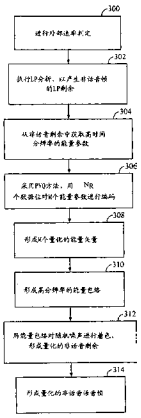
权利要求书 2 页 说明书 6 页 附图 7 页

(54)发明名称

语音中非话音部分的低数据位速率编码

(57)摘要

一种用于语音的非话音部分的低数据速率编码方案,它包括这样一些步骤:从语音帧获取高时间分辨率能量系数,使能量系数量化,从量化的能量系数中产生高时间分辨率的能量包络,以及用能量包络的量化值形成随机产生的噪声矢量从而重新构筑残留信号。能量包络可以采用线性插入技术来产生。可以获得后处理测量,并将其与预定的阈值比较,以确定编码规则是否执行恰当。



CN 1815558 B

1. 一种对非话音语音进行低数据位速率语音编码的方法，其特征在于，它包含：

将输入的语音帧标识为非话音的语音帧；

对所述非话音语音帧进行线性预告分析，以产生非话音的线性预告残余；

从所述非话音的线性预告残余中，获取高时间分辨率的能量参数；

对所述高时间分辨率的能量参数进行编码；

对所述高时间分辨率的能量参数进行量化处理，形成经量化的能量矢量；

形成高时间分辨率的能量包络；

通过用所述高时间分辨率的能量包络使随机噪声着色 (coloring)，生成量化的非话音残余；以及

生成量化的非话音语音帧，

其中，形成高分辨率能量包络包含按照下面的计算，从经解码的能量值 W_i，i = 1,2，3，…M，来形成 N 个取样高时间分辨率的能量包络 ENV[n]、语音帧的长度，这里，n = 1,2，3，…N，

M-2 个能量值代表语音当前残余的 M-2 个子帧的能量，每一子帧具有的长度是 L = N/(M-2)；

W_1 和 W_M 的值分别代表最后的残余帧过去 L 个取样的能量以及下一残余帧未来 L 个取样的能量；

W_{m-1}、W_m 和 W_{m+1} 分别代表第 (m-1)、第 m 和第 (m+1) 个子带的能量；

对于 n = m*L-L/2 至 n = m*L+L/2，代表第 m 个子帧的能量包络 ENV[n] 的取样值计算为：

对于 n = m*L-L/2 直至 n = m*L，

$$ENV[n]=\sqrt{W_{m-1}}+(1/L)*(n-m*L+L)*(\sqrt{W_m}-\sqrt{W_{m-1}})$$；以及

对于 n = m*L 直至 n = m*L+L/2，

$$ENV[n]=\sqrt{W_m}+(1/L)*(n-m*L)*(\sqrt{W_{m+1}}-\sqrt{W_m})$$，

其中，所述计算能量包络 ENV[n] 的步骤是假设 m = 2,3,4，…，M，对于 M-1 个带中的每一个，重复进行的，以计算整个能量包络 ENV[n]，这里，对于当前残余帧，n = 1,2，…,N。

2. 如权利要求 1 所述的方法，其特征在于，所述获取高时间分辨率能量参数包含获取 M 个本地能量参数 E_i，其中，i = 1,2，…,M，它是通过执行下述步骤而从非话音残余 R[n] 中获得的：

将 N 个取样残余 R[n] 划分成 (M-2) 个子块 X_i，其中，i = 2,3，…,M-1，每一子块 X_i 具有长度 L = N/(M-2)；

从前一帧过去量化残余中获取 L 个取样过去残余块 X_1；

从后一帧的线性预告残余中获取 L 个取样未来残余块 X_M；

按照下面的等式，从 M 个子块中的每一子块 X_i，i = 1,2，…，M，中生成 M 个本地能量参数 E_i，这里，i = 1,2，…,M：

$$E=\frac{1}{L}*\sum_{m=1}^{L}X_i[m]*X_i[m]$$。

3. 如权利要求 1 所述的方法，其特征在于，形成高时间分辨率能量包络包含采用从下

一帧得到的先行参数值以及从前一帧得到的前一参数值，使用于帧边界处的当前帧的能量包络光滑。

4. 如权利要求 1 所述的方法，其特征在于，对所述高时间分辨率能量参数进行编码包含按照锥形进位矢量量化方法对所述能量参数进行编码。

5. 一种对非话音语音进行低数据位速率语音编码的语音编码器，其特征在于，它包含：

将输入的语音帧标识为非话音的语音帧的装置；

对所述非话音语音帧进行线性预告分析以产生非话音的线性预告残余的装置；

从所述非话音的线性预告残余中获取高时间分辨率的能量参数的装置；

对所述高时间分辨率的能量参数进行编码的装置；

对所述高时间分辨率的能量参数进行量化处理以形成经量化的能量矢量的装置；

形成高时间分辨率的能量包络的装置；

通过用所述高时间分辨率的能量包络使随机噪声着色以生成量化的非话音残余的装置；以及

生成量化的非话音语音帧的装置，

其中，形成高分辨率能量包络包含按照下面的计算，从经解码的能量值 W_i，i = 1,2,3，…M，来形成 N 个取样高时间分辨率的能量包络 ENV[n]、语音帧的长度，这里，n = 1,2,3，…N，

M-2 个能量值代表语音当前残余的 M-2 个子帧的能量，每一子帧具有的长度是 L = N/(M-2)；

W_1 和 W_M 的值分别代表最后的残余帧过去 L 个取样的能量以及下一残余帧未来 L 个取样的能量；

W_{m-1}、W_m 和 W_{m+1} 分别代表第 (m-1)、第 m 和第 (m+1) 个子带的能量；

对于 n = m*L-L/2 至 n = m*L+L/2，代表第 m 个子帧的能量包络 ENV[n] 的取样值计算为：

对于 n = m*L-L/2 直至 n = m*L，

$$ENV[n] = \sqrt{W_{m-1}} + (1/L)*(n - m*L + L)*(\sqrt{W_m} - \sqrt{W_{m-1}})$$；以及

对于 n = m*L 直至 n = m*L+L/2，

$$ENV[n] = \sqrt{W_m} + (1/L)*(n - m*L)*(\sqrt{W_{m+1}} - \sqrt{W_m})$$，

其中，所述计算能量包络 ENV[n] 的步骤是假设 m = 2,3,4，…，M，对于 M-1 个带中的每一个，重复进行的，以计算整个能量包络 ENV[n]，这里，对于当前残余帧，n = 1,2，…，N。

语音中非话音部分的低数据位速率编码

[0001] 本申请是申请日为 1999 年 11 月 12 日、申请号为 99815573.X、发明名称为“语音中非话音部分的低数据位速率编码”发明专利申请的分案申请。

技术领域

[0002] 本发明总的涉及语音处理领域，本发明尤其涉及语音中非话音部分的低数据位速率编码的方法和装置。

背景技术

[0003] 采用数字技术进行话音传输已经非常广泛，尤其是在长途和数字无线电话应用领域更是这样。接着，这又在确定可以在信道上发送的最少信息量并同时保持重新构筑的语音感觉质量方面，引起了人们的兴趣。如果发送信息是通过简单地进行取样和数字化来进行的，则为实现传统的模拟电话语音质量时需要每秒 64 千位 (kbps) 数量级的数据速率。然而，通过采用语音分析，随后采用适当的编码、传输，再在接收机处重新合成，可以大大减小数据速率。

[0004] 我们把采用获取与人的语音发生模型有关的参数对语音进行压缩的技术的装置称为语音编码器。语音编码器将输入的语音信号分为一些时间段，或者是一些分析帧。语音编码器通常包括编码器或译码器，或编码译码器。编码器对输入的语音帧进行分析，并获取某些相关的参数，随后将这些参数量化成二进制表述，即，量化成一组数据位或二进制的数据包。这些数据包在通信信道上传送到接收机和译码器。译码器对数据包进行处理，并将它们解量化，产生参数，随后再用这些解量化的参数，对这些语音帧进行重新合成。

[0005] 语音编码器的作用是通过去除语音中所有固有的自然冗余，将数字化的语音信号压缩成低数据位速率的信号。数字压缩是通过用一组参数来代表输入的语音帧并用量化来代表具有一组数据位的参数来实现的。如果输入的语音帧的数据位数是 N_i，而由语音编码器所产生的数据包的数据位数是 N_o，那么由语音编码器所实现的压缩倍数是 $C_r = N_i/N_o$。我们所面临的挑战是在实现目标压缩倍数的同时，保持高话音质量的译码语音。语音编码器的性能取决于 (1) 上述语音模型或分析及合成处理过程的组合的良好程度，以及 (2) 在每帧的目标数据位速率 N_o 时，参数量化过程进行的量化程度。所以，语音模型的目标是用每帧较少的一组参数，来捕获语音信号的基本部分或目标话音质量。

[0006] 在低数据位速率下有效地对语音进行编码的一种有效的技术是多模式编码。多模式编码对不同类型的输入语音帧实施不同的模式规则或编译码规则。每一种模式或编译码过程以最有效的方式来表达某种类型的语音段（即，发声的、不发声的，或者是背景噪声）。采用一种外部模式决定机构来检查输入的语音帧，并对采用什么模式用于该帧作出决定。通常，通过从输入的帧中取出几个参数，并对它们进行评估，而作出采用哪一种模式的决定，以开环方式决定所采用的模式。所以，模式决定是在事先不知道输出语音的准确情况即按照语音质量或其他的特性测量来说输出语音与输入的语音有多大的相似程度而作出的。语音编译码器的一种典型的开环模式决定见美国专利 5,414,796，该专利已转让给本发

明的受让人。

[0007] 多模式编码可以是固定速率的，对每一帧采用相同数量的数据位 N_0；也可以采用变速率的，这时，不同的模式采用不同的数据位速率。变速率编码仅采用将编译码器参数编码成适合获得目标质量水平的数据位数。因此，采用变数据位速率 (VBR) 技术，在明显较低的平均速率下，可以得到与固定速率、更高速率编码器相同的目标话音质量。典型的变速率语音编码器见美国专利 5,414,796，该专利已转让给本发明的受让人。

[0008] 目前，人们无论是在商业上还是在研究兴趣上都强烈地希望开发一种能在中等的到较低数据位速率（在 2.4 到 4kbps 或以下的范围内）下工作的高质量的语音编码器。其应用范围包括无线电话、卫星通信、互联网电话、各种多媒体和话音流应用、话音邮件以及其他的话音储存系统。其驱动力是在数据包丢失的情况下，需要具有高容量，以及对较强性能的要求。近来建立各种语音编码标准的努力是推动低速语音编码规则的研究和开发的另一直接的驱动力。低速语音编码器在每一许可的应用带宽下生成更多的信道或用户，并且与合适信道编码附加层耦合的低速语音编码器可以适合编码器技术规范的整个数据位预算，并在信道出现差错的情况下，仍具有较强的性能。

[0009] 所以，多模式 VBR 语音编码是一种在低数据位速率下对语音进行编码的有效的机制。传统的多模式技术需要对各个语音段（如，非话音的、话音的以及过渡部分）设计有效的编码方案或模式以及用于背景噪声或无声的模式。语音编码器的全部性能取决于每一种模式工作的良好程度，而编码器的平均速率取决于用于非话音的、话音的、以及语音其他部分不同模式的数据位速率。为了实现低平均速率下的目标质量，必须设计一些有效的、高性能的模式，并且其中的某些模式必须在较低的数据位速率下工作。通常，话音的和非话音的语音段是在高数据速率下捕获的，而背景噪声和无声部分是用在明显较低的速率下工作的模式来代表的。所以，需要有一种低数据速率的编码技术，在采用每一帧最少数量的数据位的时候能够捕获语音的非话音部分。

[0010] 发明概述

[0011] 本发明是一种采用每一帧最少数量的数据位准确捕获语音的非话音部分的低数据速率编码技术。因此，按照本发明对语音的非话音部分进行编码的方法最好包括这样一些步骤，即，从一个语音帧中获取高时间分辨率的能量系数；对高时间分辨率的能量系数进行量化处理；从经量化的能量系数中产生高时间分辨率的能量包；并且通过使随机生成的噪声矢量具有能量包络的量化值来重新构筑剩余的信号。

[0012] 本发明还提供了一种对语音的非话音部分进行编码的语音编码器，它包括从一个帧的语音中获取高时间分辨率的能量系数的装置；使高时间分辨率的能量系数量化的装置；从量化的能量系数中产生高时间分辨率的能量包络的装置；以及通过使随机产生的噪声矢量具有量化的能量包络值来重新构筑残留信号的装置。

[0013] 本发明还提供了对语音的非话音部分进行编码的语音编码器，它最好包括从一个帧的语音中获取高时间分辨率的能量系数的模块；使高时间分辨率的能量系数量化的模块；从量化的能量系数中产生高时间分辨率的能量包络的模块；以及通过使随机产生的噪声矢量具有量化的能量包络值来重新构筑残留信号的模块。

[0014] 附图简述

[0015] 图 1 是由语音编码器在每一端处终断的通信信道的方框图。

[0016] 图 2 是一编码器的方框图。

[0017] 图 3 是一译码器的方框图。

[0018] 图 4 是描述对用于语音的非话音部分进行低数据速率编码的技术的步骤的流程图。

[0019] 图 5A-E 给出的是信号幅度对于离散时间的关系。

[0020] 图 6 是描绘锥形进位矢量量化编码过程的功能方框图。

[0021] 较佳实施例的详细描述

[0022] 图 1 中，第一编码器 10 接收数字化的语音取样 s(n)，并对取样信号 s(n) 进行编码，用于在传输介质 12 或通信信道 12 上传输到第一译码器 14。译码器 14 对经编码的语音取样信号进行译码，并合成输出语音信号 $s_{合成}(n)$。对于沿相反方向上进行的传输，第二编码器 16 对数字化的语音取样信号 s(n) 进行编码，而该取样信号是在通信信道 18 上传输的。第二译码器 20 接收经编码的语音取样信号，并对其进行译码，产生经合成的输出语音信号 $s_{合成}(n)$。

[0023] 语音取样信号 S(n) 代表已经按照本领域方法（如，脉冲编码调制 (PCM)、压扩 μ 律或 A 律）中的任何一种方法数字化和量化的语音信号。

[0024] 正如本领域中人们所知道的那样，语音取样信号 S(n) 被组织成输入数据帧，其中，每一帧包含预定数量的数字化语音取样信号 s(n)。在一种典型的实施例中，采用 8kHz 的取样速率，这时，每一 20 毫秒的帧包含 160 个取样信号。在下面描述的实施例中，从 8kbps（全速率）到 4kbps（二分之一速率）到 2kbps（四分之一速率）到 1kbps（八分之一），数据传输的速率在逐个帧的基础上是可变的。最好数据传输速率是可变的，这是因为对于包含相对较少语音信息的数据帧来说，可以有选择地采用较低的数据速率。正如本领域中的普通技术人员所了解的那样，也可以采用其他的取样速率、帧大小和数据传输速率。

[0025] 第一编码器 10 和第二译码器 20 一起包含一个第一语音编码器或语音编译码器。同样，第二编码器 16 和第一译码器 14 一起包含一个第二语音编码器。本领域中的技术人员能够理解，语音编码器能够用数字信号处理器 (DSP)、专用集成电路 (ASIC)、离散电路的逻辑门电路、固件或传统的可编程软件模块和微处理器来构成。软件模块可以做在 RAM 存储器、按块擦除存储器、寄存器、或本领域中已知的其他形式的可写储存介质。也可以用任何一种传统的处理器、控制器或状态机来代替微处理器。特别设计用于语音编码的专用集成电路见美国专利 5,727,123 和申请日为 1994 年 2 月 16 日、标题为“声码器专用集成电路”的美国专利申请 08/197,417，二者均已转让给本发明的受让人。

[0026] 图 2 中，可以用在语音编码器中的编码器 100 包括：模式决定模块 102、基音估计模块 104、LP 分析模块 106、LP 分析滤波器 108、LP 量化模块 110 和残留量化模块 112。输入语音帧 s(n) 被提供到模块决定模块 102、基音估计模块 104、LP 分析模块 106 以及 LP 分析滤波器 108。模式决定模块 102 根据每一输入语音帧 s(n) 的周期性，产生模式索引 I_M 和模式 M。按照周期性对语音帧进行分类的各种方法见申请日为 1997 年 3 月 11 日、标题是“METHOD AND APPARATUS FOR PERFORMING REDUCEDRATE VARIABLE RATE VOCODING”的美国专利申请 08/815,354，该专利申请已转让给本发明的受让人。这些方法也已并入电信行业协会行业暂行标准 TIA/EIA IS-127 和 TIA/EIA IS-733。

[0027] 基音估计模块 104 根据每一输入的语音帧 s(n) 产生基音索引 I_P 和滞后值 P_0。LP

分析模块 106 对每一输入的语音帧 s(n) 执行线性预告分析，产生 LP 参数 a。LP 参数 a 被提供到 LP 量化模块 110。LP 量化模块 110 还接收模式 M。LP 量化模块 110 产生 LP 索引 I_{LP} 以及经量化的参数 $\hat{a}$。LP 分析滤波器 108 除了输入语音帧 s(n) 以外还接收经量化的 LP 参数 $\hat{a}$。LP 分析滤波器 108 产生 LP 残留信号 R[n]，它代表输入语音帧 s(n) 和量化的线性预告参数 $\hat{a}$ 之间的误差。LP 残留 R[n]、模式 M 和量化 LP 参数 $\hat{a}$ 被提供到残留量化模块 112。根据这些值，残留量化模块 112 产生残留索引 I_R 和经量化的残留信号 $\hat{R}[n]$。

[0028] 图 3 中，语音编码器中可以使用的译码器 200 包括 LP 参数译码模块 202、剩余译码模块 204、模式译码模块 206 以及 LP 合成滤波器 208。模式译码模块 206 接收模式索引 I_M 并对其进行译码，由此产生模式 M。LP 参数译码模块 202 接收模式 M，和 LP 索引 I_{LP}。LP 参数译码模块 202 对接收值进行译码，以产生经量化的 LP 参数 $\hat{a}$。剩余译码模块 204 接收剩余索引 I_R、基音索引 I_P 和模式索引 I_M。剩余译码模块 204 对接收值进行译码，产生量化的残留信号 $\hat{R}[n]$。经量化的残留信号 $\hat{R}[n]$ 和经量化的 LP 参数 $\hat{a}$ 被提供到 LP 合成滤波器 208，由它来合成经译码的输出语音信号 $\hat{S}[n]$。

[0029] 图 2 所示编码器 100 各种模块的操作和构成以及图 3 中所示译码器是本领域中已知的，其详细描述见 L.B Rabiner 和 R.W.Schafer 的 Digital Processing ofSpeech Signal，396-453(1978)。典型的编码器和典型的译码器见美国专利 5,414,796。

[0030] 图 4 中的流程图描述了一种按照一种实施例用于语音的非话音段低数据速率编码技术。图 4 中所示的低速率非话音编码模式提供了一种在更低平均数据速率下的多模式语音编码器，通过准确捕获每一帧数量较少的数据位的非话音部分，它保留了整体较高的话音质量。

[0031] 在步骤 300，编码器对非话音的以及不是非话音的输入语音帧执行外部数量确定和识别。速率的确定是通过考虑到从语音帧 S[n] 获取的几个参数来完成的，这里，n = 1，2，3，…，N，比如，帧的能量 (E)、帧的周期 (Rp) 以及频谱倾斜 (Ts)。将这些参数与一组预定的阈值比较。根据比较的结果，判断当前帧是否是非话音的。如下所述，如果当前帧是非话音的，则将其编码为非话音的帧。

[0032] 按照下面的等式，可以确定帧的能量：

[0033] $$E = \frac{1}{N} * \sum_{m=1}^{N} S[m] * S[m]$$

[0034] 按照下面的等式，可以决定帧的周期：

[0035] Rp = 所有 k 中的最大值{$\Re$ (S[n],S[n+k])}，k = 1,2，…，N

[0036] 这里，$\Re$ (x[n],x[n+k])是 x 的自相关函数。按照下面的等式，可以确定频谱倾斜：

[0037] Ts = (Eh/El)

[0038] 这里，Eh 和 El 是 Sl[n] 和 Sh[n] 的能量值，Sl 和 Sh 是原始语音帧 S[n] 的低通和高通分量，它们可以由一组低通滤波器和高通滤波器来产生。

[0039] 在步骤 302，进行 LP 分析，产生非话音帧的线性预告剩余。线性预告 (LP) 是采用本领域中众所周知的技术来完成的，详见美国专利 5,414,796，和 L.B.Rabiner 与 R.W.Schafer 的 Digital Processing of Speech Signals 396-458(1978)。N 取样的非话

音 LP 剩余 R[n] 是从输入语音帧 S[n] 中产生的，这里，n = 1,2，…，N。正如在上面对比文献中所描述的那样，采用已知的 LSP 量化技术，在线性频谱对 (LSP) 域中使 LP 参数量化。原始语音信号幅度与离散时间索引之间的关系见图 5A 中所示。经量化的非话音语音信号幅度与离散时间索引之间的关系见图 5B 所示。原始非话音剩余信号幅度与离散时间索引之间的关系见图 5C 所示。能量包络幅度与离散时间索引之间的关系见图 5D 所示。经量化的非话音残留信号幅度与离散时间索引之间的关系见图 5E 所示。

[0040] 在步骤 304，获取非话音残留信号的精细时间分辨率能量参数。执行下面的步骤，从非话音剩余 R[n] 中获取几个 (M) 本地能量参数 E_i，这里，i = 1,2，…，M。将 N 个取样剩余 R[n] 分成 (M-2) 子块 X_i，这里，i = 2,3，…，M-1，每一块 X_i 的长度是 L = N/(M-2)。从前一帧的过去 (past) 量化剩余中得到 L 个取样的过去剩余块 X_1。(L 个取样的过去剩余块 X_1 含有最后语音帧 N 个取样剩余的最后 L 个取样)。从下一个帧的 LP 剩余中得到 L 个取样的将来剩余块 X_M。(L 个取样的将来剩余块 X_M 含有下一个语音帧 N 取样 LP 剩余开头的 L 个取样。) 按照下面的等式，从 M 个块 X_i 中的每一个中产生 M 个本地能量参数 E_i，这里，i = 1,2，…，M。

[0041] $$E = \frac{1}{L} * \sum_{m=1}^{L} X_i[m] * X_i[m]$$

[0042] 在步骤 306，按照锥形进位矢量量化 (PVQ) 方法，用 Nr 个数据位，对 M 个能量参数进行编码。所以，用 Nr 个数据位对 M-1 个本地能量值 E_i 进行编码，形成量化的能量值 W_i，这里，i = 2,3，…，M。采用数据位 N_1, N_2，…，N_K 的 K 个步骤的 PVQ 编码方案，从而 $N_1+N_2+\cdots+N_K = Nr$，即，用于量化非话音剩余 R[n] 的数据位总数。对于 k 个级 (stage) 中的每一个级，执行下面的步骤 (这里，k = 1,2，…，K)。对于第一级 (即，k = 1)，将频带数设置在 $B_k = B_1 = 1$，并且频带长度设置在 $L_k = 1$。对于每一频带 B_k，按照下面的等式，设置平均值 $mean_j$，这里，j = 1,2，…，B_k：

[0043] $$mean_j = \frac{1}{L_j} * \sum_{m=1}^{L_j} E_m$$

[0044] 用 $N_k = N_1$ 将 B_k 平均值 $mean_j$ 量化，而形成平均值 $qmean_j$ 的量化组，这里，j = 1,2，…，B_k。将属于每一频带 B_k 的能量除以相关量化的平均值 $qmean_j$，而产生新的一组能量值 $\{E_{k,i}\} = \{E_{1,i}\}$，这里，i = 1,2，…，M。在第一级的情况下 (即，对于 k = 1)，对于每一 i，(i = 1,2，…，M)：

[0045] $E_{1,I} = E_i/qmeans_1$

[0046] 分成子频带、获取每一频带的平均值、用每一级的数据位使平均值量化，并且随后将子频带的分量除以子带的量化平均值，对于每一以后的级 k，重复这一过程，这里 k = 2,3，…，K-1。

[0047] 在第 k 级，采用全部 N_k 个数据位，用为每一频带而设计的各个 VQ，使 B_k 子频带中每一个的分矢量量化。M = 8 以及级 = 4 的 PVQ 编码过程是通过图 6 中所示的例子来描述的。

[0048] 在步骤 308，形成 M 个量化的能量矢量。通过用最终剩余的分矢量和量化平均值最终使上述 PVQ 编码过程反向，从编码簿 (codebook) 和代表 PVQ 信息的 Nr 个数据位中形成 M 个量化的能量矢量。图 7 中通过举例，描述了 M = 3 以及级 k = 3 时的 PVQ 译码过程。

正如本领域中的普通技术人员能够理解的那样，非话音的 (UV) 增益可以用任何一种传统的编码技术来量化。编码技术方案并非仅限于图 4-7 中所描述的实施例的 PVQ 方案。

[0049] 在步骤 310，形成高分辨率的能量包络。按照下面计算，从经译码的能量值 W_i，形成 N 个取样（即，语音帧的长度），高时间分辨率的能量包络 ENV[n]，这里，n = 1,2,3, …,N, i = 1,2,3, …, M。M-2 个能量值代表语音当前剩余 M-2 个子帧的能量，每一子帧的长度 L = N/(M-2)。W_1 和 W_M 的值分别代表最后的剩余帧的过去的 L 个取样，和下一个剩余帧未来 L 个取样的能量。

[0050] 如果 W_{m-1}、W_m 和 W_{m+1} 分别代表第 m-1 个、第 m 个和第 m+1 个子带的能量，那么对于 n = m*L-L/2 至 n = m*L+L/2，代表第 m 个子帧的能量包络 ENV[n] 的采样计算如下：对于 n = m*L-L/2，一直到 n = m*L，

[0051] $$ENV[n]=\sqrt{W_{m-1}}+(1/L)*(n-m*L+L)*(\sqrt{W_m}-\sqrt{W_{m-1}})$$

[0052] 并且对于 n = m*L，一直到 n = m*L+L/2，

[0053] $$ENV[n]=\sqrt{W_m}+(1/L)*(n-m*L)*(\sqrt{W_{m+1}}-\sqrt{W_m})$$

[0054] 假设 m = 2,3,4, …,M，对于 M-1 个频带中的每一个频带，重复对能量包络 ENV[n] 进行计算的步骤，以计算整个能量包络 ENV[n]，这里，对于当前剩余帧，n = 1,2, …, N。

[0055] 在步骤 312，通过使能量包络 ENV[n] 对随机噪声进行着色，形成量化后的非

[0056] 话音残留信号。按照下面的等式，形成量化后的非话音剩余 qR[n]：

[0057] qR[n] =噪声 [n]*ENV[n]，n = 1,2, …, N

[0058] 这里，噪声 [n] 是具有单位方差的随机白噪声信号，它是由与编码器和译码器同步的随机数发生器模拟产生的。

[0059] 在步骤 314，形成量化的非话音语音帧。正如在本领域中以及在上述美国专利 5,414,796 中以及 L.B.Rabiner 与 R.W.Schafer 在 Digital Processing of SpeechSignal, 396-458(1978) 中所描述的那样，采用传统的 LP 合成技术，通过将量化后的非话音语音进行逆向 LP 滤波，产生量化的非话音剩余 qS[n]。

[0060] 在一种实施例中，通过测量感测的 (perceptual)) 误差测量如感测的信噪比 (PSNR)，可以执行质量控制步骤，而 PSNR 定义如下：

[0061] $$PSNR=10*\log 10\frac{\sum_{n=1}^{N}(x[n]-e[n])^2}{\sum_{n=1}^{N}e[n]*e[n]}$$

[0062] 这里，x[n] = h[n]*R[n]，而 e(n) = h[n]*qR[n]，“*”表示卷积或滤波操作，h(n) 是感测的加权 LP 滤波器，而 R[n] 和 qR[n] 分别是原始的和量化的非话音剩余。将 PSNR 与一预定的阈值比较。如果 PSNR 小于该阈值，则非话音编码方案就不会进行恰当地得到执行，并且可以执行更高速率的编码方式，代替更精确地捕获当前帧。另一方面，如果 PSNR 超过预定的阈值，则非话音的编码方案就得到了很好的执行，并保留该模式判断。

[0063] 上文中已经描述了本发明的较佳实施例。然而，对本领域中普通技术人员而言，在不偏离本发明的精神和范围的情况下，还可以对这些实施例作各种各样的修正。所以，本发明并非仅限于这些实施例，而应当以权利要求书来限定本发明。

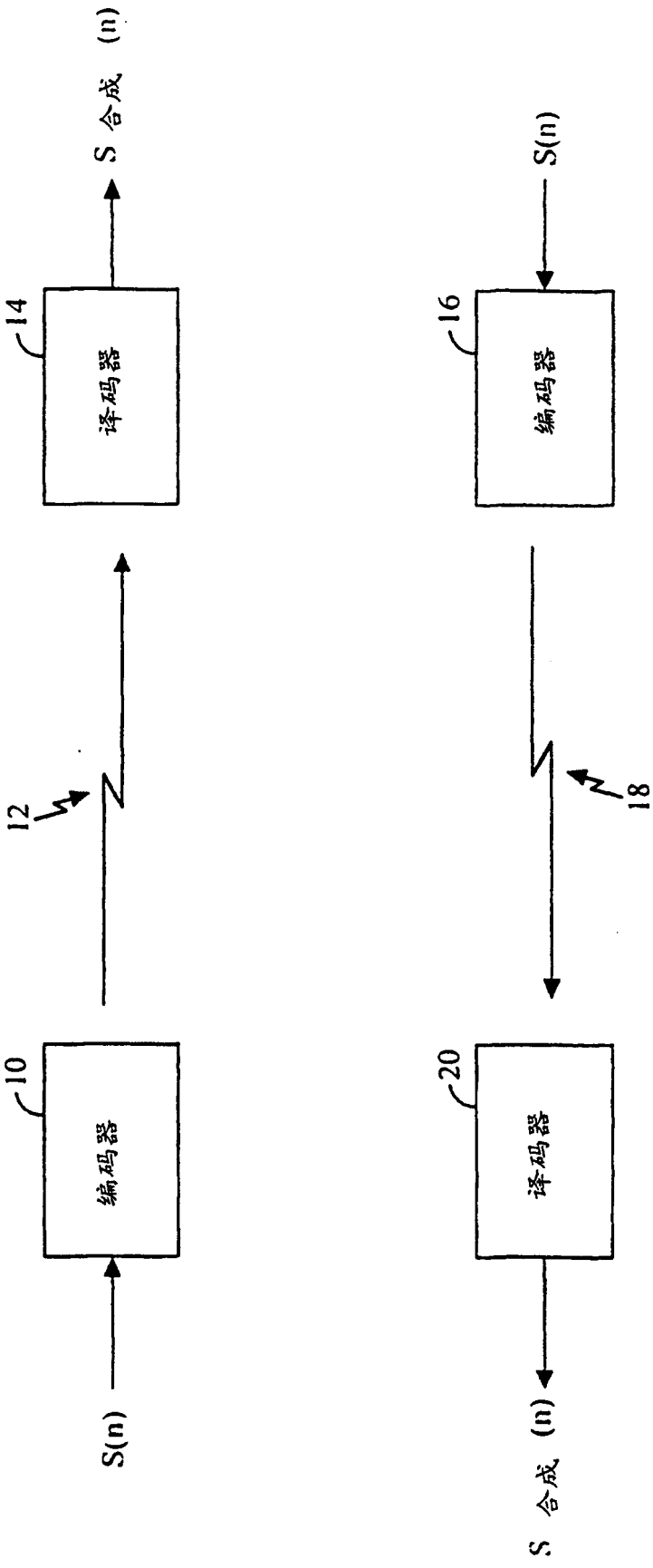


图 1

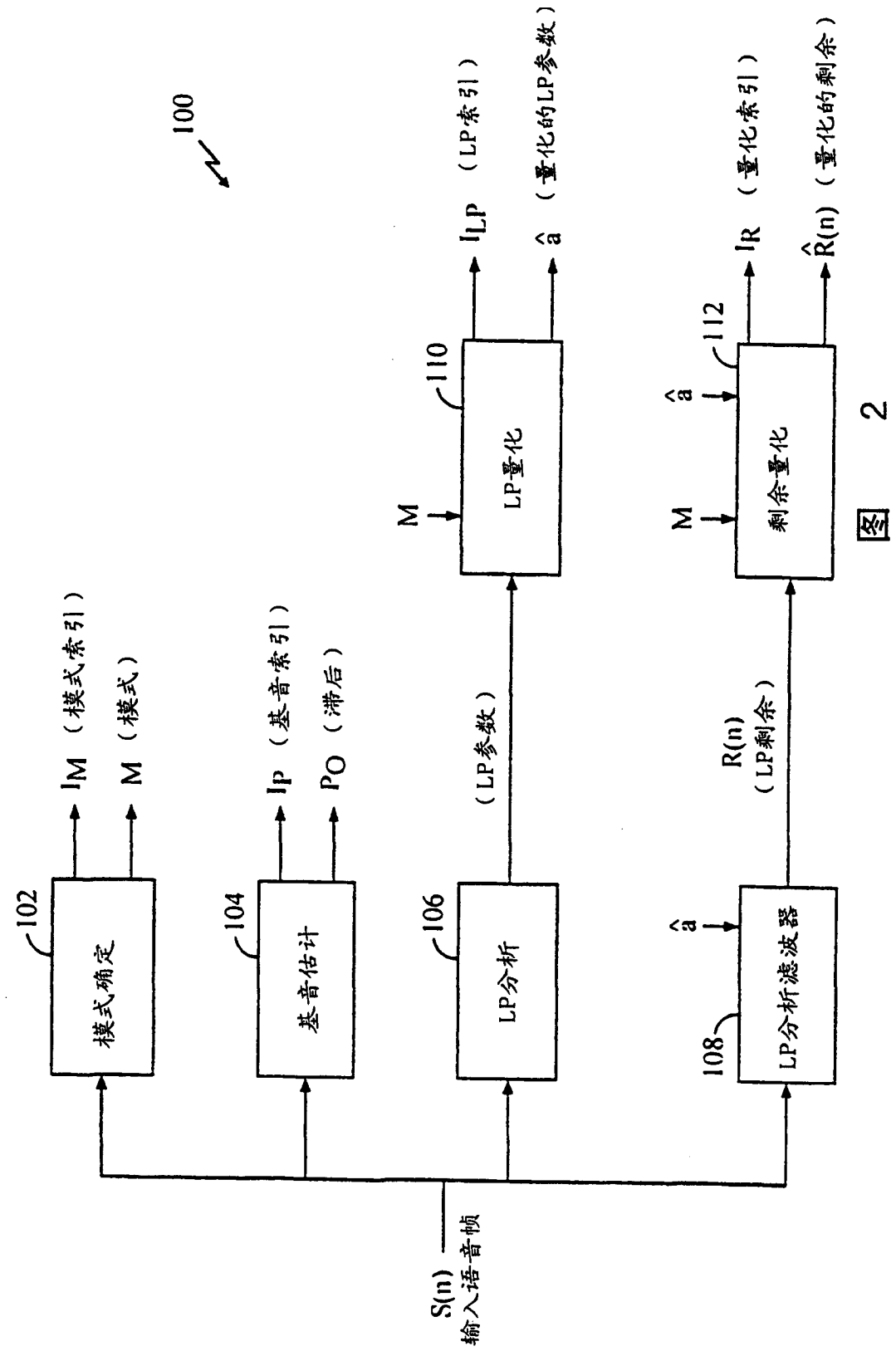

100
S(n)
输入语音帧
102
模式确定
I_M（模式索引）
M（模式）
104
基音估计
I_P（基音索引）
PO（滞后）
106
LP分析
（LP参数）
108
LP分析滤波器
$\hat{a}$
R(n)
（LP剩余）
110
LP量化
M
I_{LP}（LP索引）
$\hat{a}$（量化的LP参数）
112
剩余量化
M
$\hat{a}$
I_R（量化索引）
$\hat{R}(n)$（量化的剩余）

图 2

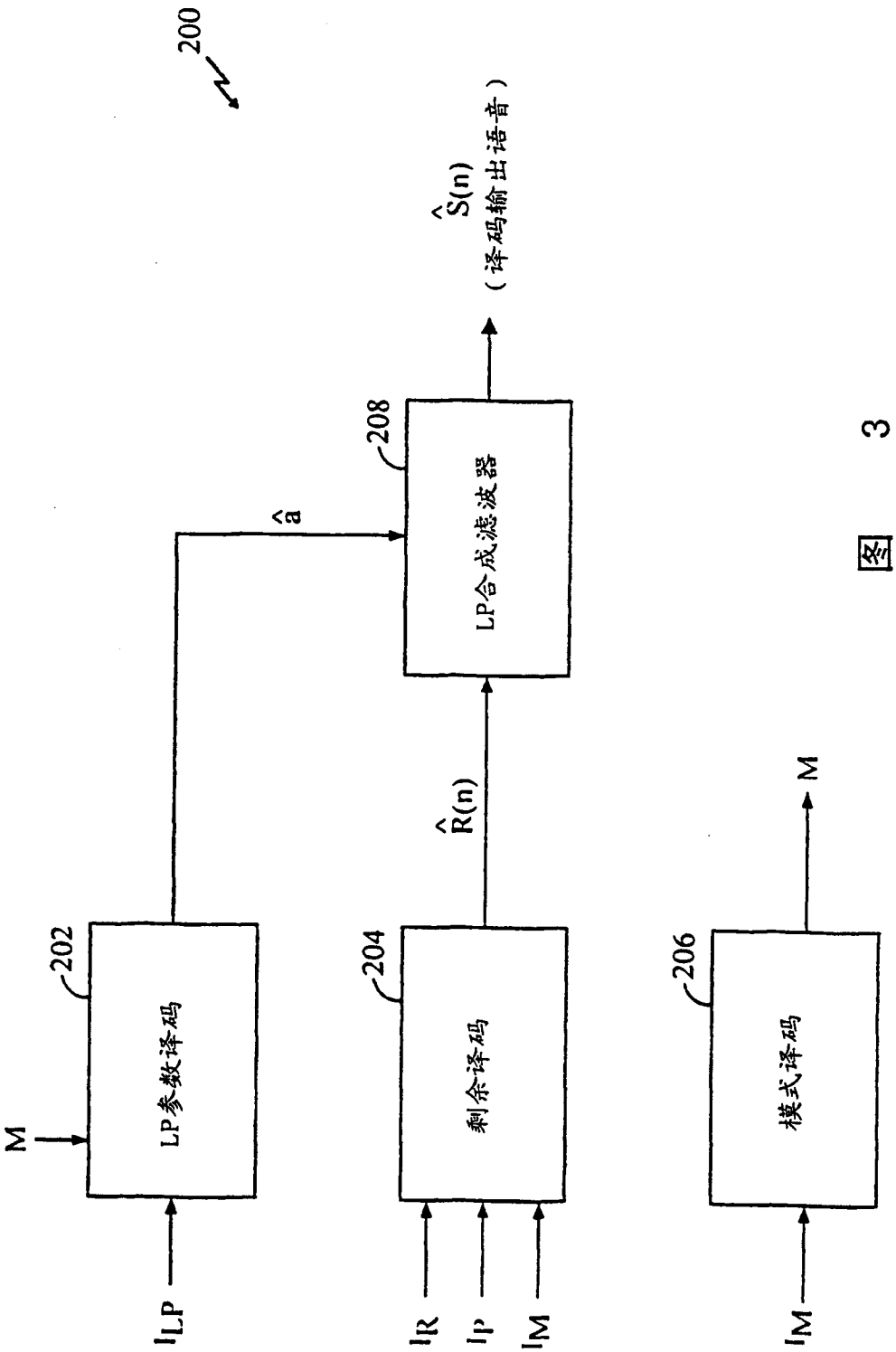

200
$\hat{S}(n)$
（译码输出语音）
208
LP合成滤波器
$\hat{a}$
$\hat{R}(n)$
202
LP参数译码
M
I_{LP}
204
剩余译码
I_R
I_P
I_M
206
模式译码
M
I_M

图 3

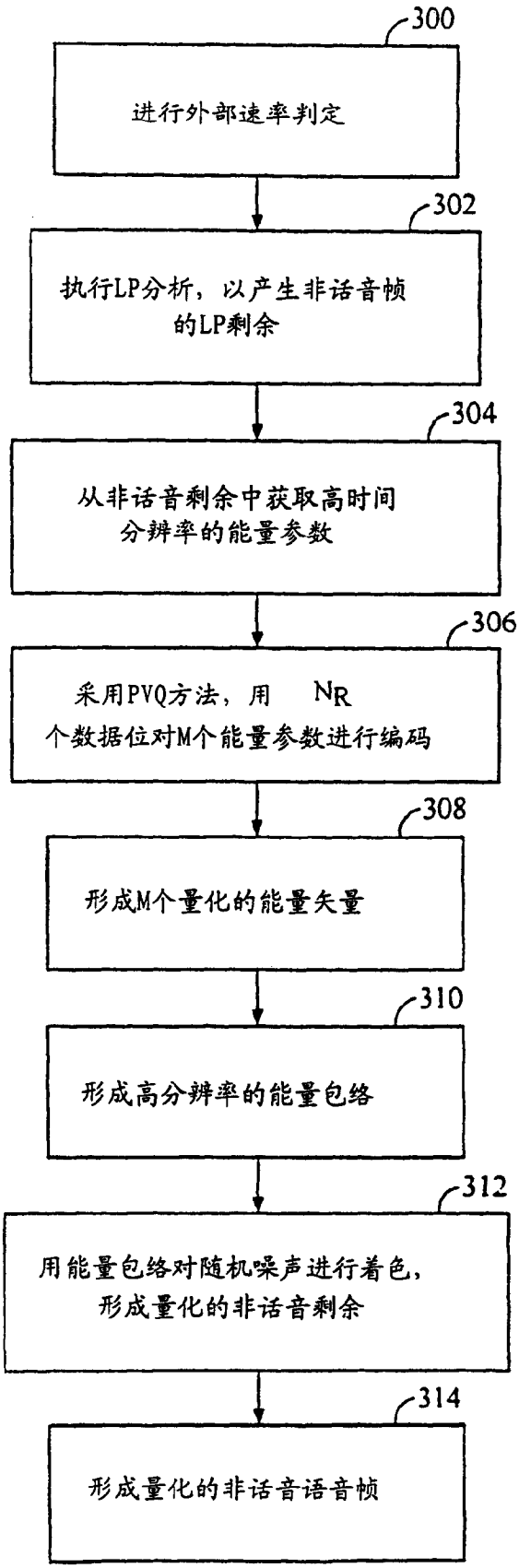
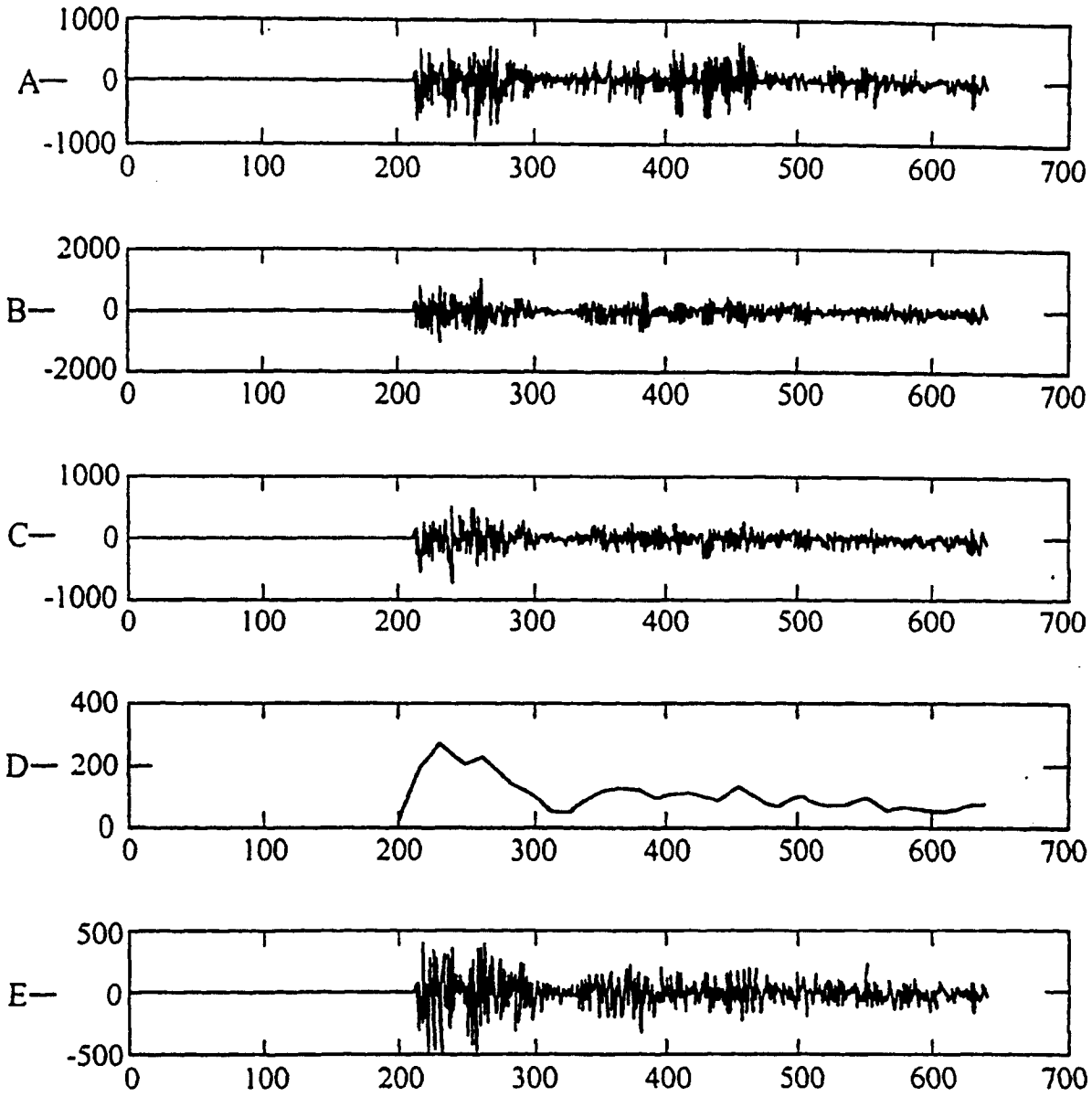

300
进行外部速率判定
302
执行LP分析，以产生非话音帧的LP剩余
304
从非话音剩余中获取高时间分辨率的能量参数
306
采用PVQ方法，用 N_R 个数据位对M个能量参数进行编码
308
形成M个量化的能量矢量
310
形成高分辨率的能量包络
312
用能量包络对随机噪声进行着色，形成量化的非话音剩余
314
形成量化的非话音语音帧

图 4



A	原始非话音语音	S_{UV}
B	量化语音	S_{UV-Q}
C	原始非话音剩余	$R_{UV}[n]$
D	能量包络	$ENV[n]$
E	量化剩余	$R_{UV-Q}[n]$

图 5

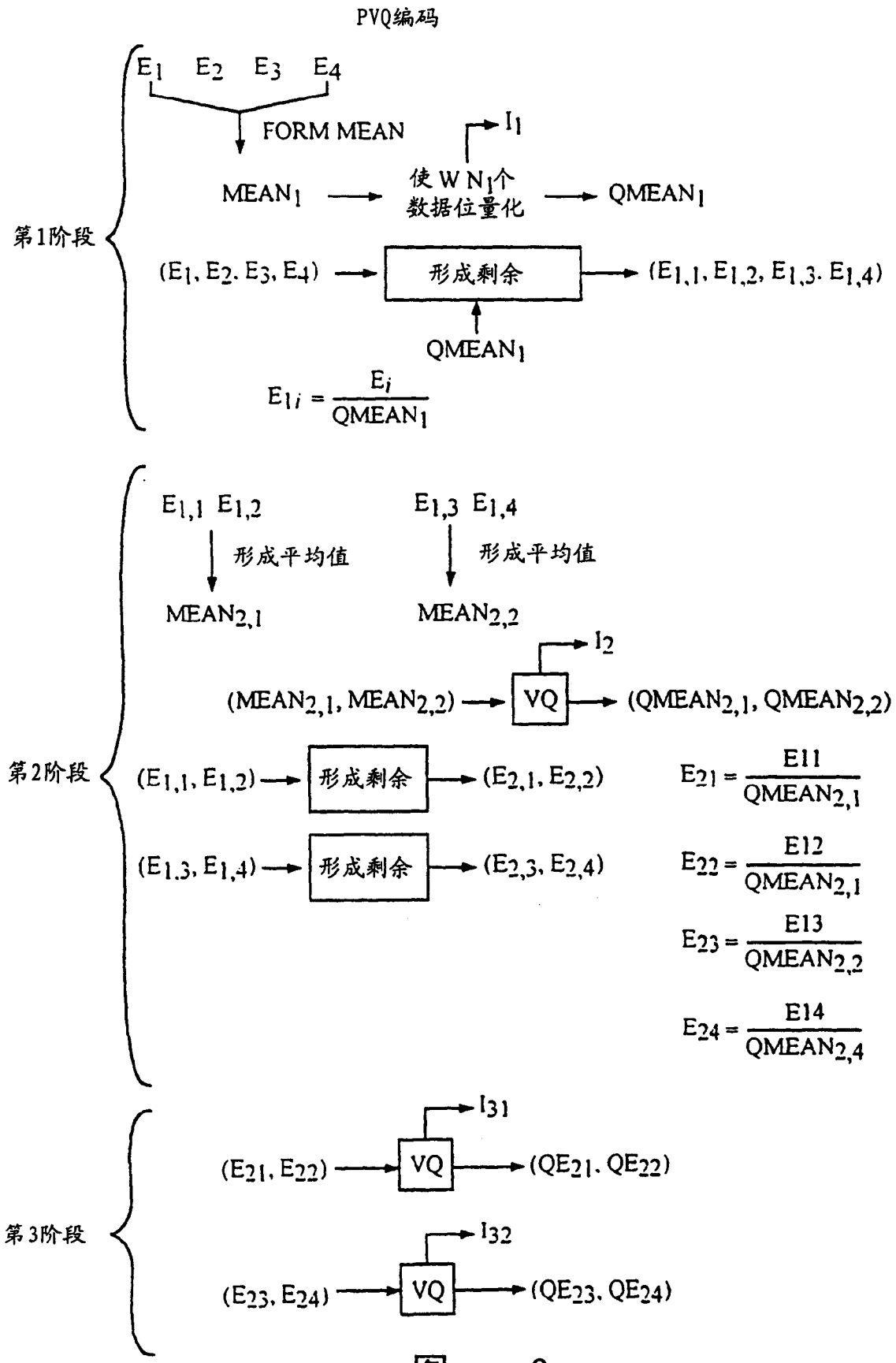
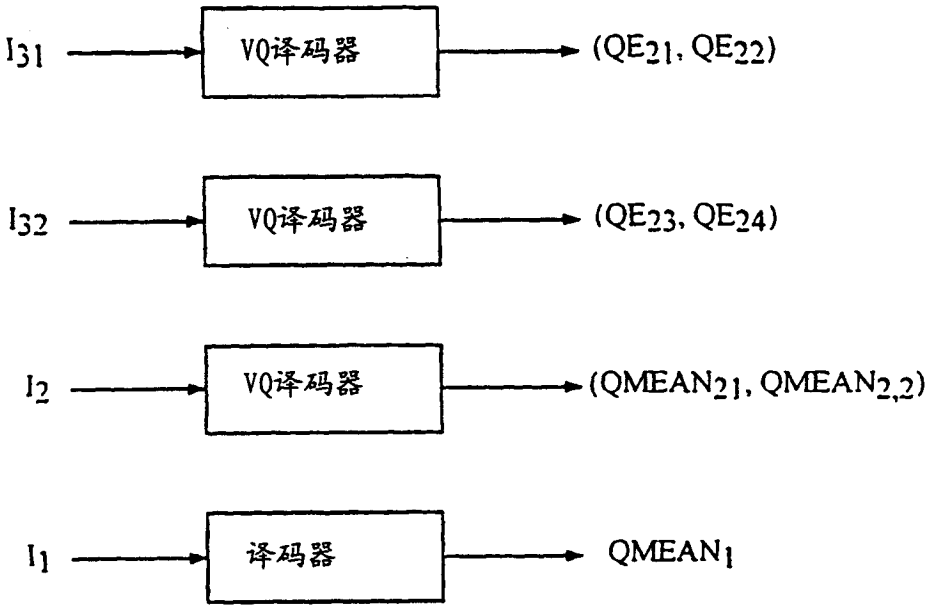

PVQ编码
第1阶段
E1 E2 E3 E4
FORM MEAN
I1
MEAN1
使W N1个数据位量化
QMEAN1
(E1, E2. E3, E4)
形成剩余
(E1,1, E1,2, E1,3. E1,4)
QMEAN1
E1i = Ei / QMEAN1
第2阶段
E1,1 E1,2
形成平均值
MEAN2,1
E1,3 E1,4
形成平均值
MEAN2,2
I2
(MEAN2,1, MEAN2,2)
VQ
(QMEAN2,1, QMEAN2,2)
(E1,1, E1,2)
形成剩余
(E2,1, E2,2)
E21 = E11 / QMEAN2,1
(E1,3, E1,4)
形成剩余
(E2,3, E2,4)
E22 = E12 / QMEAN2,1
E23 = E13 / QMEAN2,2
E24 = E14 / QMEAN2,4
第3阶段
I31
(E21, E22)
VQ
(QE21, QE22)
I32
(E23, E24)
VQ
(QE23, QE24)

图 6

PVQ编码



使用的总数据位 $N_0 = I_1 + I_2 + I_{31} + I_{32}$.

译码能量值:

$$W_1 = QE_1 = QMEAN_1 * QMEAN_{2,1} * QE_{21}$$
$$W_2 = QE_2 = QMEAN_1 * QMEAN_{2,1} * QE_{22}$$
$$W_3 = QE_3 = QMEAN_1 * QMEAN_{2,2} * QE_{23}$$
$$W_4 = QE_4 = QMEAN_1 * QMEAN_{2,2} * QE_{24}$$

图 7