

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.  
G10L 19/02 (2006.01)



# [12] 发明专利申请公布说明书

[21] 申请号 200780015598.X

[43] 公开日 2009年5月13日

[11] 公开号 CN 101432804A

[22] 申请日 2007.3.12

[21] 申请号 200780015598.X

[30] 优先权

[32] 2006.3.13 [33] FR [31] 0602179

[86] 国际申请 PCT/FR2007/050915 2007.3.12

[87] 国际公布 WO2007/104889 法 2007.9.20

[85] 进入国家阶段日期 2008.10.30

[71] 申请人 法国电信公司

地址 法国巴黎

[72] 发明人 P·菲利普 C·沃 P·科郎

[74] 专利代理机构 中国国际贸易促进委员会专利  
商标事务所  
代理人 郭思宇

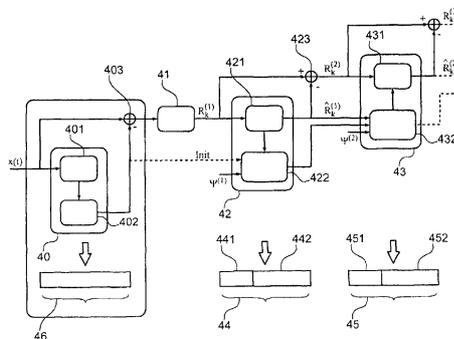
权利要求书 3 页 说明书 18 页 附图 3 页

## [54] 发明名称

对源音频信号进行编码的方法和相应的编码设备、解码方法和设备、信号、计算机程序产品

## [57] 摘要

本发明涉及对源音频信号进行编码的方法。按照本发明，这种方法包括下列步骤：按照至少两种不同的编码技术对表示源音频信号的至少一个变换的系数的量化分布进行编码，给出至少两个表示一个量化分布的数据组；根据预定选择准则选择表示量化分布的所述数据组中的一个数据组；以及发送和/或存储表示所选择的量化分布的数据组和表示相应编码技术的指示符。



1. 一种对源音频信号进行编码的方法，其特征在于包括下列步骤：

按照至少两个不同的编码技术对表示所述源音频信号的至少一个变换的系数的量化分布进行编码，给出表示量化分布的至少两个数据组；

按照基于分别从所述数据组重建的信号的失真的度量和基于对所述数据组进行编码所需的比特率的选择准则，选择表示量化分布的所述数据组中的一个数据组；以及

发送和/或存储表示所选择的量化分布的所述数据组和表示相应的编码技术的指示符。

2. 按照权利要求 1 所述的编码方法，其特征在于：对于至少所述编码技术中的第一编码技术，所述数据组与所述量化分布的参数表示相应。

3. 按照权利要求 2 所述的编码方法，其特征在于：所述参数表示由至少一段由斜率和原点值表征的直线形成。

4. 按照权利要求 1 至 3 中任一项所述的编码方法，其特征在于：所述编码技术中的第二编码技术给出恒定的量化分布。

5. 按照权利要求 1 至 4 中任一项所述的编码方法，其特征在于：按照第三编码技术，所述量化分布与绝对听觉门限相应。

6. 按照权利要求 1 至 5 中任一项所述的编码方法，其特征在于：按照第四编码技术，所述表示量化分布的数据组包括所有所实施的量化间隔。

7. 按照权利要求 1 至 6 中任一项所述的编码方法，其特征在于：所述编码实现分级处理，给出包括一个初级和至少一个细化级的至少两个分级编码级，所述细化级包括对所述初级或前一细化级的细化信息。

8. 按照权利要求 7 所述的编码方法，其特征在于：按照第五编

码技术，表示量化分布的所述数据组在给定细化级通过考虑前一分级编码级所构建的数据得出。

9. 按照权利要求 7 和 8 中任一项所述的编码方法，其特征在于：所述选择步骤在每个分级编码级执行。

10. 按照权利要求 1 至 9 中任一项所述的编码方法，其特征在于：所述方法给出一些系数帧，为每个帧执行所述选择步骤。

11. 一种对源音频信号进行编码的设备，其特征在于包括：

按照至少两个不同的编码技术对表示所述源音频信号的至少一个变换的系数的量化分布进行编码、给出表示量化分布的至少两个数据组的装置；

按照基于分别从所述数据组重建的信号的失真的度量和基于对所述数据组进行编码所需的比特率的选择准则选择表示量化分布的所述数据组中的一个数据组的装置；以及

发送和/或存储表示所选择的量化分布的所述数据组和表示相应的编码技术的指示符的装置。

12. 一种可从通信网络下载和/或存储在计算机可读载体内和/或可由微处理器执行的计算机程序产品，其特征在于包括实现按照权利要求 1 至 10 中任一项所述的编码方法的程序代码指令。

13. 一种表示源音频信号的经编码的信号，包括表示量化分布的数据，其特征在于包括：

表示在编码时根据选择准则从至少两个可用技术中选出的一个对所实现的量化分布进行编码的技术的指示符，所述选择准则基于分别从按照所述技术编码的量化分布重建的信号的失真的度量和基于按照所述技术对量化分布进行编码所需的比特率；以及

表示相应量化分布的数据组。

14. 按照权利要求 13 所述的信号，其特征在于包括：关于通过分级处理得到的包括一个初级和至少一个细化级的至少两个分级的数据，所述细化级包括对所述初级或前一细化级的细化信息；以及表示每个所述级的编码技术的指示符。

15. 按照权利要求 13 和 14 中任一项所述的信号，其特征在于：所述信号被组织在一些相继的系数的帧中，所述信号包括表示用于每个所述帧的编码技术的指示符。

16. 一种对表示源音频信号的包括表示量化分布的数据的经编码信号进行解码的方法，其特征在于包括下列步骤：

从所述经编码信号提取：

表示在编码时根据选择准则从至少两个可用技术中选出的一个对所实现的量化分布进行编码的技术的指示符，所述选择准则基于分别从按照所述技术编码的量化分布重建的信号的失真的度量和基于按照所述技术对量化分布进行编码所需的比特率，以及

表示所述相应量化分布的数据组；以及

根据所述数据组和由所述指示符标明的编码技术重建所述重建的量化分布。

17. 按照权利要求 16 所述的解码方法，其特征在于包括考虑所述重建的量化分布构建表示所述源音频信号的重建音频信号的步骤。

18. 一种对表示源音频信号的包括表示量化分布的数据的经编码信号进行解码的设备，其特征在于包括：

从所述经编码信号提取下列各项的装置

表示在编码时根据选择准则从至少两个可用技术中选出的一个对所实现的量化分布进行编码的技术的指示符，所述选择准则基于分别从按照所述技术编码的量化分布重建的信号的失真的度量和基于按照所述技术对量化分布进行编码所需的比特率，以及

表示所述相应量化分布的数据组；以及

根据所述数据组和由所述指示符标明的编码技术重建所述重建的量化分布的装置。

19. 一种可从通信网络下载和/或存储在计算机可读载体内和/或可由微处理器执行的计算机程序产品，其特征在于包括实现按照权利要求 16 至 17 中任一项所述的解码方法的程序代码指令。

## 对源音频信号进行编码的方法和相应的编码设备、 解码方法和设备、信号、计算机程序产品

### 技术领域

本发明涉及对诸如音乐之类的音频数字信号或数字化语音信号进行编码和解码的技术领域。

具体地说，本发明涉及在实现感知编码（perceptual encoding）中量化音频信号的谱系数的技术。

本发明特别（但不只是）可应用于在 MPEG 音频标准（ISO/IEC 14496-3）的环境内所提出的使用可扩展数据编码/解码型系统对音频数字数据进行分级编码（hierarchical encoding）的系统。

一般地说，本发明可以应用于对声音和音乐进行高效量化以便存储、压缩和通过传输信道（例如，无线或有线信道）传输声音和音乐的技术领域。

### 背景技术

#### 1 传输掩蔽曲线的感知编码

##### 1.1 音频压缩和量化

音频压缩往往基于人耳的一定听觉能力。对音频信号的编码和量化往往要考虑这个特性。在这种情况下所用的术语是“感知编码”，或者遵从人耳心理听觉模型的编码。

人耳不能区分一个信号在邻近频率和在有限时隙内发送的两个分量。这个性质称为听觉掩蔽（auditory masking）。此外，耳朵具有听觉门限（hearing threshold），在安静的环境内，将感觉不到所发送的低于听觉门限的声音。这个门限的值随声波频率改变。

在压缩和/或传输音频数字信号中，可以想到的是，确定量化比特数，以量化形成信号的频谱分量，不会引入过多的量化噪声而损害

到编码后信号的质量。目标通常是减少量化比特数量以得到对信号的高效压缩。因此，必须做的是找到声音质量与信号压缩程度之间的折衷。

所以，在经典的现有技术中，量化原理使用了人耳引起的掩蔽门限和掩蔽性质，以确定注入音频信号而在信号再现时耳朵感觉不到（即不会引入任何过分失真）的可接受的最大量化噪声量。

## 1.2 感知音频变换编码

音频变换编码的详细说明可参见 Jayant、Johnson 和 Safranek 的“基于人类感知的方法的信号压缩”（“Signal Compression Based on Method of Human Perception”，Proc. Of IEEE, Vol. 81, No. 10, pp. 1385-1422, October 1993）。

这种技术使用了图 1 所例示的人耳频率掩蔽模型，图中示出了音频信号频率与人耳掩蔽门限之间的关系例子。X 轴 10 表示频率  $f$  (Hz)，Y 轴 11 表示声强  $I$  (dB)。人耳将信号  $x(t)$  的频谱在频域内按 Bark 刻度分解成一些关键频带 (critical band) 120、121、122、123。信号  $x(t)$  的关键频带  $n$  120 具有能量  $E_n$ ，在频带  $n$  内和在邻近的关键频带 122 和 123 处产生掩蔽 13。所关联的掩蔽门限 13 与“掩蔽”分量 120 的能量  $E_n$  成正比，而随关键频带低于和高于关键频带  $n$  递减。

分量 122 和 123 在图 1 这个例子中受到掩蔽。此外，分量 121 由于低于绝对听觉门限 14 也受到掩蔽。于是，通过将绝对听觉门限 14 和与音频信号  $x(t)$  的以关键频带分析的各个分量关联的掩蔽门限组合在一起，就可以得到总的掩蔽曲线。这个掩蔽曲线表示可以在信号编码时叠加到信号上而人耳感觉不到的最大量化噪声的谱密度。于是，在量化从源音频信号频率变换得到的谱系数期间，就可形成量化间隔分布 (quantization interval profile)，不很严格地也可称为注入噪声分布。

图 2 为例示经典的感知编码器的原理的流程图。时间源音频信号

$x(t)$ 由时间-频率变换单元 20 变换到频域。于是,得到由谱系数 $X_n$ 形成的信号的频谱。该频谱由心理听觉模型 21 分析,该模型根据绝对听觉门限和信号的每个频谱分量的掩蔽门限确定信号的总掩蔽曲线 C。所得到的掩蔽曲线可用来得出可以注入的量化噪声量,从而确定用来量化谱系数或样本的比特数。确定比特数这个步骤由给出每个谱系数 $X_n$ 的量化间隔分布 $\Delta_n$ 的二进制分配单元 22 执行。二进制分配单元通过用掩蔽曲线 C 所给出的成形限制修改量化间隔设法达到目标比特率。量化间隔 $\Delta_n$ 由二进制分配单元 22 以比例因子 F 的形式编码,再在比特流 T 内作为辅助信息发送。

量化单元 23 接收到谱系数 $X_n$ 和所确定的量化间隔 $\Delta_n$ 后给出经量化的系数 $\hat{X}_n$ 。

最后,编码和比特流形成单元 24 将经量化的谱系数 $\hat{X}_n$ 和比例因子 F 集中在一起进行编码,形成含有经编码的源音频信号的有效负载数据和表示比例因子的数据的比特流。

## 2 掩蔽曲线的分级构建

下面说明在音频数字数据分级编码环境内现有技术的缺点。然而,本发明可以应用于根据人耳心理听觉模型实现量化的各种音频数字信号编码器。这些编码器不一定是分级的。

分级编码需要将一些编码级级联起来。第一级产生比特率最低的编码版本,而随后的这些级提供逐级增大比特率的相继改善。在音频信号编码的特定情况下,这些改善级传统上基于如在以上这节中所说明的感知变换编码。

然而,这种分级感知变换编码的一个缺点是必须从最前面的一级或者说初级就发送所得到的比例因子。这些比例因子与有效负载数据相比占用了分配给低比特率级的一大部分比特率。

为了克服这个缺点以节约对注入量化噪声分布(即比例因子)的传输,J. Li 在“采用隐含听觉掩蔽的嵌入式音频编码(EAC)” (“Embedded Audio Coding (EAC) With Implicit Auditory

Masking”，ACM Multimedia 2002）中业已提出一种称为“隐含”技术的掩蔽技术。这种技术有赖于分级结构的编码/解码系统，在每个改善级用掩蔽曲线的近似值对掩蔽曲线进行递推估计，使掩蔽曲线逐级细化。

因此，在分级编码的每个级利用在前一级所量化的变换系数重复更新掩蔽曲线。

由于对掩蔽曲线的估计基于时间-频率变换的系数的量化值，因此能在编码器和解码器处同样执行，这具有避免向解码器传输量化间隔分布或量化噪声的优点。

### 3 现有技术的缺点

即使基于分级编码的隐含掩蔽技术避免了传输掩蔽曲线因此与经典的需发送量化间隔分布的感知编码相比在比特率上有所得益，但本发明的发明者注意到它还具有一些缺点。

确实，在编码器和解码器内同时实现的掩蔽模型一定是闭环的（closed-ended），因此可以不用精确地适合信号的特性。例如，无论需编码的频谱分量有没有音调特性都使用单个掩蔽因子。

此外，掩蔽曲线是在信号是标准（standing）信号的假设下计算的，不能适当地应用于瞬变部分和声冲击。

此外，由于掩蔽曲线是在每一级从前一级所量化的系数或系数残差得出，因此第一级的掩蔽曲线由于某些部分的频谱还没有编码所以是不完全的。这个不完全的曲线不一定表示所考虑的这个分级编码级的量化间隔分布的最佳形状。

### 发明内容

本发明提出了一种对源音频信号进行编码的方法，这种方法包括下列步骤：

按照至少两个不同的编码技术对表示源音频信号的至少一个变换的系数的量化分布进行编码，给出表示量化分布的至少两个数据组；

按照基于分别从所述数据组重建的信号的失真的度量和基于对所述数据组进行编码所需的比特率的选择准则，选择表示量化分布的所述数据组中的一个数据组；以及

发送和/或存储表示所选择的量化分布的所述数据组和表示相应的编码技术的指示符。

因此，本发明所依赖的是一种新颖的、创造性的对源音频信号的系数进行编码的途径，这种途径可以降低分配给传输量化间隔的比特率，同时还将注入量化噪声分布保持成尽可能接近从对信号的完整知识计算得出的掩蔽曲线所给定的分布。

本发明提出在计算量化间隔分布的不同可行模式之间进行选择。因此，可以在若干量化间隔分布或注入噪声分布的模板之间进行选择。这个选择由指示符，例如包含在由编码器形成的发送给音频信号再现系统即解码器的比特流内的信号，报告。

选种准则可以主要考虑每个量化分布的效率和对应数据组进行编码所需的比特率。

因此，在传送表示信号的数据所需的比特率与影响信号的失真之间得到折衷。

量化于是得到优化。同时，使得发送不直接提供音频信号本身的表示量化间隔分布的信息的数据所需的比特率最小。

也就是说，在解码器处，量化模式的选择通过将根据需编码的音频信号估计的基准掩蔽曲线与分别与各个量化模式关联的噪声分布相比较来实现。

与现有技术相比，本发明的技术改善了压缩效率，因此提供了更好的感知质量。

对于编码技术中至少第一编码技术来说，数据组可以与量化分布的参数表示相应。

也就是说，在所提出的量化经变换的音频信号的系数的这些技术中，有以参数表示量化分布的可能性。

在一个特定实施例中，参数表示由至少一个由斜率和初始值表征

的直线段形成。

第二编码技术可以给出恒定的量化分布。

这种编码模式因此提出根据信噪比(SNR)而不是根据信号的掩蔽曲线对量化间隔分布编码。

按照第三有益编码技术,量化分布与绝对听觉门限相应。

也就是说,表示量化分布的数据组可以是空的,编码器不用向解码器发送任何量化分布数据。绝对听觉门限对于解码器来说是已知的。

按照第四编码技术,表示量化分布的数据组可以包括所有所实施的量化间隔。

这种第四编码技术对应于量化间隔分布根据只有编码器知道的信号的掩蔽曲线确定、完全发送给解码器的情况。所需的比特率高,但信号的再现质量最佳。

在一个特定实施例中,编码实现分级处理,给出包括一个初级和至少一个细化级的至少两个分级编码级,所述细化级包括对初级或前一细化级的细化信息。

在这种情况下,采用第五编码技术,表示量化分布的数据组在给定的细化级通过考虑前一分级编码级所构建的数据得出。

因此本发明可以高效地应用于分级编码,提出按照在每个分级编码级细化量化间隔分布的技术对量化间隔分布进行编码。

选择步骤可以在每个分级编码级执行。

如果编码方法给出一些系数帧,可以为每个帧执行选择步骤。

因此,不仅可以为每个处理帧执行信号传送,而且在对数据分级编码的特定应用中可以为每个细化级执行信号传送。

在其他情况下,编码可以对包括一些具有预定或可变长度的帧的组执行。还可以规定,只要没有发送新的指示符,当前分布保持不变。

本发明还涉及包括实现这样的方法的装置的对源音频信号进行编码的设备。

本发明还涉及实现如以上所说明的编码方法的计算机程序产品。

本发明还涉及表示源音频信号的包括表示量化分布的数据的经

编码信号。这样的信号主要包括：

表示在编码时根据选择准则从至少两个可用技术中选出的一个对所实现的量化分布进行编码的技术的指示符，所述选择准则基于分别从按照所述技术编码的量化分布重建的信号的失真的度量和基于按照所述技术对量化分布进行编码所需的比特率；以及

表示相应量化分布的数据组。

这样的信号可以主要包括：关于通过分级处理得到的包括一个初级和至少一个细化级的至少两个分级编码级的数据，所述细化级包括对初级或前一细化级的细化信息；以及表示每级的编码技术的指示符。

在将本发明的信号组织在一些相继的系数的帧中时，信号可以包括表示用于每个帧的编码技术的指示符。

本发明还涉及对这样的信号进行解码的方法。这种方法主要包括下列步骤：

从经编码信号提取

表示在编码时根据选择准则从至少两个可用技术中选出的一个对所实现的量化分布进行编码的技术的指示符，所述选择准则基于分别从按照所述技术编码的量化分布重建的信号的失真的度量和基于按照所述技术对量化分布进行编码所需的比特率，以及

表示所述相应量化分布的数据组；以及

根据所述数据组和由所述指示符标明的编码技术重建所述重建的量化分布。

这种类型的解码方法还包括考虑所重建的量化分布构建表示源音频信号的重建音频信号的步骤。

对于这些编码技术中至少第一编码技术，数据组可以与量化分布的参数表示相应，而重建步骤给出以至少一个直线段的形式重建的量化分布。

对于这些编码技术中的至少第二编码技术，数据组可以是空的，

而重建步骤给出恒定的量化分布。

对于这些编码技术中的至少第三编码技术，数据组可以是空的，而量化分布与绝对听觉门限相应。

对于这些编码技术中的至少第四编码技术，数据组可以包括在以上所说明的编码方法期间所实施的所有量化间隔，而构建步骤给出呈现为在编码方法期间实施的一组量化间隔形式的量化值。

在一个特定实施例中，解码方法可以实现分级处理，给出包括一个初级和至少一个细化级的至少两个分级编码级，所述细化级包括对初级或前一细化级的细化信息。

对于这些编码技术中的至少第五编码技术，重建步骤在给定细化级给出考虑前一分级编码级所构建的数据而得到的量化分布。

本发明还涉及对表示源音频信号的经编码信号进行解码的设备，包括实现以上所说明的解码方法的装置。

本发明还涉及实现如以上所说明的解码方法的计算机程序产品。

### 附图说明

从以下以例示性和非穷举性例子的方式给出的对具体实施例的说明中和从附图中可以看到本发明的实施例的其他特征和优点，在这些附图中：

图 1 例示了频率掩蔽门限；

图 2 为按照现有技术实现的感知变换编码的简化流程图；

图 3 例示了按照本发明的信号的一个例子；

图 4 为按照本发明的编码方法的简化流程图；

图 5 为按照本发明的解码方法的简化流程图；以及

图 6A 和 6B 示意性地例示了实现本发明的编码设备和解码设备。

### 具体实施方式

#### 1 编码器结构

下面将对本发明在分级编码的具体应用中的实施例进行说明。可

以回想一下，在这种方案中，分级编码在需编码的源音频信号的时间-频率变换（例如，经修改的离散余弦变换 MDCT）的输出端确立了级联的感知量化间隔。

下面将参考图 4 对按照本发明的这个实施例的编码器进行说明。源音频信号  $x(t)$  被直接或间接地变换到频域。确实，任选地，可以首先在编码步骤 40 将信号  $x(t)$  编码。这种类型的步骤由“核心”编码器实现。在这种情况下，第一编码步骤与第一分级编码级即初始级相应。这种类型的“核心”编码器可以实现编码步骤 401 和本地解码步骤 402。于是，它给出表示以最低精细度的编码的音频信号的数据的第一比特流 46。可以设想用各种编码技术来得到这个低比特率级，例如用参量编码方案，诸如在 B. den Brinker、E. 和 W. Schuijers Oomen 的“高质量音频参量编码”（“Parametric coding for high quality audio”，in Proc. 112th AES Convention, Munich, Germany, 2002）中所揭示的正弦编码，在 M. Schroeder 和 B. Atal 的“码激线性预测（CELP）：极低比特率的高质量语音”（“Code-excited linear prediction (CELP): high quality speech at very low bit rates”，in Proc. IEEE Int. Conf. Acoust, Speech Signal Processing, Tampa, pp. 937-940 1985）中所揭示的 CELP 型分析-综合编码。

将经本地解码器 402 解码的样本与  $x(t)$  的实际值相减（步骤 403），得到时域内的残差信号  $r(t)$ 。

然后，在步骤 41，将低比特率编码器 40（或者说“核心”编码器）输出的这个残差信号从时间空间变换到频率空间。得到频域内的谱系数  $R_k^{(1)}$ 。这些系数表示“核心”编码器 40 给出的第一分级编码级的各个关键频带  $k$  的残差。

下一个编码级 42 含有对残差  $R_k^{(1)}$  编码的步骤 421，它与负责确定第一细化级的第一掩蔽曲线的心理听觉模型的实现 422 相关联。于是，在编码步骤 421 的输出端得到经量化的残差系数  $\hat{R}_k^{(1)}$ ，再将它从来自

核心编码步骤 40 的原始系数  $R_k^{(1)}$  中减去 (步骤 423)。在下一级 43 的编码步骤 431 得到新的系数  $R_k^{(2)}$  并对其进行量化和编码。在这里也实施心理听觉模型 432, 根据先前量化的残差的系数  $\hat{R}_k^{(1)}$  更新掩蔽门限。

简明地说, 基本编码步骤 40 (“核心”编码器) 可以在一个终端处传输音频信号的一个低比特率版本和对这个版本解码。在变换域内对残差进行量化的后继级 42、43 构成了一些能构建一个从低比特率级到所希望的最大比特率的分级比特流的改善级。

按照本发明, 如图 4 所示, 指示符  $\psi^{(1)}$ 、 $\psi^{(2)}$  分别与相应量化级的编码级的心理听觉模型 422、432 关联。这个指示符的值对于每个量化级来说是特定的, 控制对量化间隔分布的计算的模式。它作为经量化的谱系数的 442、452 的帧的标题 441 和 451 安置在每个经改善的编码级 42、43 所形成的所关联的比特流 44、45 内。

图 3 例示了按照这种编码技术得到的信号的结构例子。信号组织成一系列各包括标题 32 和数据字段 33 的数据块或数据帧 31。一个数据块例如与一个预定时隙的一个分级编码级的数据 (包含在数据段 33 内的) 相应。标题 32 可以包括一些有助于传信、解码等的信息片。按照本发明, 它至少包括信息  $\psi$ 。

## 2 解码器结构

下面将参考图 5 说明在对图 3 的信号进行分级解码的情况下按照本发明实现的解码方法。

以与参考图 4 给出的编码方法类似的方式, 解码包括若干解码细化级 50、51、52。

第一解码步骤 501 接收含有表示第一级在第一编码步骤期间确定的发送给解码器的指示符  $\psi^{(1)}$  的数据 530 的比特流 53。这个比特流还含有表示音频信号的谱系数的数据 531。

按照所接收的这些经量化的系数或经量化的系数残差和  $\psi^{(1)}$  的值, 在第一级 502 实施心理听觉模型, 以确定掩蔽曲线的第一估计, 从而确定量化间隔分布, 供在解码方法的这一级处的解码器用来处理

可得到的谱系数残差。

所得到的每个关键频带  $k$  的谱系数残差  $\hat{R}_k^{(1)}$  使得在步骤 512 可以更新下一级 51 的心理听觉模型，于是细化掩蔽曲线，从而细化量化间隔分布。因此，这个细化考虑了级 2 的包含在相应编码器所发送的比特流 54 的标题 540 内的指示符  $\psi^{(2)}$  的值、前一级的量化残差和与包含在比特流 54 内的与级 2 有关的经量化的数据 541。

在第二解码级 51 的输出端得到经量化的残差  $\hat{R}_k^{(2)}$ 。残差  $\hat{R}_k^{(2)}$  与前一级的残差  $\hat{R}_k^{(1)}$  相加 (56)，此外还注入下一级 52，类似，级 52 改善从解码步骤 51 和在步骤 522 的心理听觉模型的实现得到的谱系数和量化间隔分布的精度。这一级还接收编码器发送的含有指示符  $\psi^{(3)}$  的值 55 和经量化的频谱 551 的比特流 55。

所得到的经量化的残差  $\hat{R}_k^{(3)}$  与残差  $\hat{R}_k^{(2)}$  相加，诸如此类推。

总而言之，心理听觉模型随着系数被相继的细化级解码而更新。读出编码器所发送的指示符  $\psi$ ，每个量化级就可以重建噪声分布（或量化分布）。

下面将详细说明对于按照具体实施例的编码方法和解码方法来说是共同的对心理听觉模型和谱系数量化模型进行更新的步骤。然后，详细说明确定编码时所执行的指示符  $\psi$  的值的步骤，再说明在解码器内重建量化间隔的步骤。

### 3 心理听觉模型更新

可以回想一下，心理听觉模型考虑了音频信号被人耳分解成的子频带，因此利用心理听觉信息可以确定掩蔽门限。这些门限用来确定谱系数的量化间隔。

在本发明中，心理听觉模型更新掩蔽曲线的步骤（在编码方法的步骤 422、432 和在解码方法的步骤 502、512、522 实现）在选择量化间隔分布上无论指示符  $\psi$  的值如何都保持不变。

相反，心理听觉模型使用所更新的掩蔽曲线的方式却由指示符  $\psi$  的值而定，以确定量化谱系数（或在前一细化级所确定的残差系数）

所需的量化间隔分布。

在每个量化级 $l$ （在分级编码-解码系统的这个具体应用中），心理听觉模型使用所估计的音频信号 $x(t)$ 的频谱 $\hat{X}_k^{(l)}$ ，其中 $k$ 表示时间-频率变换的频率附标。这个频谱在第一量化细化级用在核心编码器所实现的编码步骤的输出端可得到的数据初始化。在随后的量化级，根据在上一细化级的输出端的所量化的残差系数 $\hat{R}_k^{(l-1)}$ 按照式 $\hat{X}_k^{(l)} = \hat{X}_k^{(l-1)} + \hat{R}_k^{(l-1)}$ ， $k = 0, \dots, N - 1$ 更新频谱 $\hat{X}_k^{(l)}$ ，其中 $N$ 为变换在频域内的长度。

通过将频谱 $\hat{X}_k^{(l)}$ 与心理听觉模型所得到的掩蔽模式进行卷积，就可以重建与信号 $x(t)$ 关联的掩蔽门限。

于是，得到在量化步骤 $l$ 所估计的掩蔽曲线 $\hat{M}_k^{(l)}$ ，作为与信号 $x(t)$ 关联的掩蔽门限和绝对听觉曲线中最大的那个值。

此外，编码和解码步骤各包括在首次根据核心编码器发送的数据实施心理听觉模型（编码方法的步骤 422 和解码方法的步骤 502）期间对心理听觉模型初始化的步骤 *Init*。

可以根据所实现的核心编码器的类型设想若干方案，在附录中揭示了其中的一些例子。

#### 4 量化谱系数

在精确说明确定决定对量化分布的选择的指示符 $\psi$ 的最佳值的技术前，首先详细说明本发明的在得知量化间隔分布后计算需分配给量化音频信号的各个谱系数的比特数的方式。

##### 4.1 二进制分配

在这里所说明的是量化定律 $Q$ 的一般情况，例如可以与将值四舍五入为最接近的整数相应。输入量化级 $l$ 的残差系数 $R_k^{(l)}$ 的经量化值 $\hat{R}_k^{(l)}$ 按照下式根据量化间隔分布 $\Delta_n^{(l)}$ 得出：

$$rq_k^{(l)} = Q \left( g_l \frac{R_k^{(l)}}{\Delta_n^{(l)}} \right) \quad \text{对于 } kOffset(n) \leq k \leq kOffset(n+1)$$

和

$$\hat{R}_k^{(l)} = \frac{\Delta_n^{(l)}}{g_l} Q^{-1}(rq_k^{(l)}) \quad \text{对于 } kOffset(n) \leq k \leq kOffset(n+1)$$

其中  $rq_k^{(l)}$  是有整数值的系数，而  $kOffset(n)$  为关键频带  $n$  的起始频率附标。

这部分的系数  $g_l$  相当于可以调整与由  $\Delta_n^{(l)}$  给出的分布并行注入的量化噪声电平的恒定增益。

在第一途径中，增益  $g_l$  由分配环路确定，以便达到分配给各个量化级  $l$  的目标比特率。然后，增益  $g_l$  在量化级的输出端的比特流内发送给解码器。

在第二途径中，增益  $g_l$  只是细化级  $l$  的函数，而这个函数对于解码器来说是已知的。

#### 4.2 量化间隔分布

于是，本发明的编码和解码方法建议根据在若干编码技术或分布计算模式之间的选择确定量化间隔分布  $\Delta_n^{(l)}$ 。该选择由在比特流内发送的指示符  $\psi$  的值表示。根据这个指示符的值，或者全部发送或者部分发送甚至完全不发送量化间隔分布。在这种情况下，量化间隔分布在解码器内估计。

量化级  $l$  所用的量化间隔分布  $\Delta_n^{(l)}$  根据本级可得到的掩蔽曲线和根据输入端的指示符  $\psi^{(l)}$  计算。

在一个具体实施例中，指示符  $\psi^{(l)}$  编码成 3 个比特，以表示 5 种不同的对量化间隔分布编码的技术。

对于指示符  $\psi^{(l)}$  的值 = 0 的情况，不使用心理听觉模型所估计的掩蔽曲线，量化间隔分布是均匀的，遵从式  $\Delta_n^{(l)} = \text{cte}$ 。也就是说，在信

噪比 (SNR) 意义上进行量化。

对于指示符 $\psi^{(l)}$ 的值=1的情况, 量化间隔分布只根据绝对听觉门限按照式 $\Delta_n^{(l)} = \sum_{k=kOffset(n)}^{kOffset(n+1)-1} Q_k$ 给出, 其中 $Q_k$ 为绝对听觉门限。

在这种情况下, 编码器不向解码器发送任何量化间隔的信息。

对于指示符 $\psi^{(l)}$ 的值=2的情况, 是在级 $l$ 用心理听觉模型估计的掩蔽曲线 $\hat{M}_k^{(l)}$ 按照式 $\Delta_n^{(l)} = \sum_{k=kOffset(n)}^{kOffset(n+1)-1} \hat{M}_k^{(l)}$ 给出量化间隔分布。可以注意到的是, 这种模式只是在音频信号编码-解码系统内实现分级构建掩蔽曲线的特定应用中才是可行的。

对于指示符 $\psi^{(l)}$ 的值=3的情况, 量化间隔的分布根据可参量化和解码器已知的原型曲线给出。按照一个特定而非排它的应用, 这个原型对于各个关键频带 $n$ 是以 dB 计的仿射直线, 斜率为 $\alpha$ 。可以将 $D_n(\alpha)$ 写成:  $\log_2(D_n(\alpha)) = \alpha n + K$ , 其中 $K$ 为常数。

通过与编码器根据对需编码的信号的谱分析计算得出的基准掩蔽曲线相关, 选择斜率 $\alpha$ 的值。然后, 将它的量化值 $\hat{\alpha}$ 发送给解码器, 用来按照式 $\Delta_n^{(l)} = D_n(\hat{\alpha})$ 给出量化间隔分布。

最后, 对于指示符 $\psi^{(l)}$ 的值=4的情况, 将编码步骤所确定的量化间隔分布 $\Delta_n^{(l)}$ 完全发送给解码器。这些间隔值例如从编码器根据需编码的源音频信号计算得出的基准掩蔽曲线 $M_k$ 得出。于是, 就有:

$$\Delta_n^{(l)} = \sum_{k=kOffset(n)}^{kOffset(n+1)-1} M_k.$$

### 5 确定指示符 $\psi$ 的值

本发明提出了一种明智地选择指示符的值从而也就是选择需用来对音频信号编码和解码的量化间隔分布的具体技术。在每个量化级 $l$  (在分级编码的情况下) 的编码步骤都进行这种选择。

确实, 众所周知, 在给定的量化级, 就在需编码的信号与所重建的信号之间所感知的失真来说最佳的量化间隔分布可以通过计算基于

心理听觉模型和由式  $\Delta_n^{(l)} = \sum_{k=kOffset(n)}^{kOffset(n+1)-1} M_k^{(l)}$  给出的基准掩蔽曲线得到。选择指示符  $\psi$  的值在于在就所感知的失真来说量化间隔分布的优化与使分配给发送量化间隔分布的比特率最小之间寻求最有效的折衷。

为了得到这种类型的折衷引入成本函数

$$C(\psi) = d(\Delta_n^{(l)}(\psi), \Delta_n^{(l)}(\psi = 4)) + \theta(\psi)$$

其中,  $\psi = 0, 1, 2, 3, 4$ 。

这个函数用来考虑对量化间隔分布编码的各种技术的效率。

第一项  $d(\Delta_n^{(l)}(\psi), \Delta_n^{(l)}(\psi = 4))$  为与指示符  $\psi$  的每个所考虑的值 ( $\psi = 0, 1, 2, 3, 4$ ) 关联的量化间隔分布与最佳分布 (与指示符  $\psi$  的值 = 4 关联, 相当于基准掩蔽曲线的传输) 之间的距离的度量。可以测量这个距离, 作为以比特计的与使用“亚最佳”掩蔽分布关联的过分成本。这个成本函数按照下式计算:

$$d(\Delta_n^{(l)}(\psi), \Delta_n^{(l)}(\psi = 4)) = \sum_n \left| \log_2(\Delta_n^{(l)}(\psi)) - \log_2(\Delta_n^{(l)}(\psi = 4)) - \log_2\left(\frac{G_1}{G_2}\right) \right|$$

其中:  $G_1 = \sum_n \Delta_n^{(l)}(\psi)$ , 而  $G_2 = \sum_n \Delta_n^{(l)}(\psi = 4)$ 。

增益  $G_1$  与  $G_2$  之比可用来对量化间隔分布相互标准化。

第二项  $\theta(\psi)$  表示与量化间隔分布  $\Delta_n^{(l)}(\psi)$  的发送关联的以比特计的过分成本。也就是说, 它表示必须发送给解码器以便重建量化间隔的附加比特数 (除了对指示符  $\psi$  编码的以外)。也就是说:

对于  $\psi = 0, 1, 2$  (分别与在解码步骤期间所重新评估的对恒定量化、绝对听觉门限和掩蔽曲线编码的技术相应) 的情况,  $\theta(\psi)$  为零;

在  $\psi = 3$  时 (与对量化间隔分布进行参量编码的技术相应),

$\theta(\psi)$ 表示对 $\alpha$ 编码的比特数；以及

在 $\psi = 4$ 时（与编码器将量化间隔完全发送给解码器相应），

$\theta(\psi)$ 为根据基准曲线给出的量化间隔 $\Delta_n^{(l)}$ 进行编码的比特数。

## 6 解码方法期间量化间隔的重建

量化间隔分布的重建在量化级 $l$ 根据解码器所发送的数据执行。

无论选来对量化间隔编码的是什么技术，即无论指示符 $\psi^{(l)}$ 的值是什么，解码器首先对作为接收到的比特流的每个帧的标题所给出的这个指示符的值进行解码，再读出调整增益 $g_l$ 的值。然后，按照指示符的值区别对待，情况如下：

如果 $\psi^{(l)} = 4$ ，解码器读出全部的量化间隔 $\Delta_n^{(l)}$ ；

如果 $\psi^{(l)} = 3$ ，读出 $\alpha$ ，再在解码器按照先前所引入的式子 $\Delta_n^{(l)} = D_n(\alpha)$ 计算量化间隔分布；

如果 $\psi^{(l)} = 2$ ，解码器根据在这级 $l$ 重建的掩蔽曲线 $\widehat{M}_k^{(l)}$ 按照先前所引入的式子 $\Delta_n^{(l)} = \sum_{k=kOffset(n)}^{kOffset(n+1)-1} \widehat{M}_k^{(l)}$ 计算量化间隔的分布（递推构建）；

如果 $\psi^{(l)} = 1$ ，解码器按照先前所引入的基于绝对听觉门限的式子 $\Delta_n^{(l)} = \sum_{k=kOffset(n)}^{kOffset(n+1)-1} Q_k$ 计算量化间隔分布；以及

如果 $\psi^{(l)} = 0$ ，解码器按照先前所引入的式子 $\Delta_n^{(l)} = cte$ 计算量化间隔分布。

一旦在解码步骤计算出这些量化间隔、解码出在比特流内发送的先前引入的系数 $rq_k^{(l)}$ （相对频谱系数的有效负载数据或它们的残差值），就可以按照在相对于二进制分配的本说明书第 5.1 节中所引入的式子得到级 $l$ 的残差系数的量化值 $\widehat{R}_k^{(l)}$ 。

## 7 实现设备

本发明的方法可以用结构如图 6A 所示的编码设备实现。

这样的设备包括存储器 M 600、配有例如微处理器和由计算机程序 Pg 602 驱动的处理单元 601。初始化时，计算机程序 602 的代码指令例如装入 RAM 后由处理单元 601 的处理器执行。输入时，处理单元 601 接收需编码的源音频信号 603。处理单元 601 的微处理器  $\mu\text{P}$  按照程序 Pg 602 的指令实现以上所说明的编码方法。处理单元 601 输出比特流 604，它包括表示经编码的源音频信号的经专门量化的数据、表示量化间隔分布的数据和表示指示符  $\psi$  的数据。

本发明还提出了一种按照本发明的对表示源音频信号的编码的信号进行解码的设备，图 6B 示意性地例示了这种设备的原理性结构。这种设备包括存储器 M 610，和配有例如微处理器并由计算机程序 Pg 612 驱动的处理单元 611。初始化时，计算机程序 612 的代码指令例如装入 RAM 后由处理单元 611 的处理器执行。输入时，处理单元 611 接收比特流 613，它包括表示经编码的源音频信号的数据、表示量化间隔分布的数据和表示指示符  $\psi$  的数据。处理单元 601 的微处理器  $\mu\text{P}$  按照程序 Pg 612 的指令实现解码方法，给出所重建的音频信号 612。

## 附录

心理听觉模型可以以若干方式初始化，这取决于在初级编码步骤所实现的“核心”编码器。

### 1 根据正弦编码器发送的参数初始化

正弦编码器将音频信号模型化为具有时变频率和振幅的一系列正弦波之和。频率和振幅的量化值发送给解码器。从这些值，可以构建信号的正弦分量的频谱 $\hat{X}_k^{(0)}$ 。

### 2 根据 CELP 编码器发送的参数初始化

根据由 CELP（码激线性预测）编码器量化和发送的 LPC（线性预测编码）系数 $a_m$ ，可以按照下式得出包络频谱：

$$\hat{X}_k^{(0)} = \frac{1}{\left| 1 - \sum_{m=1}^P a_m \exp\left(-j \frac{2\pi mk}{N}\right) \right|^2}$$

其中， $N$ 为变换的长度，而 $P$ 为由 CELP 编码器发送的 LPC 系数的个数。

### 3 根据在核心编码器输出端解码的信号初始化

初始频谱 $\hat{X}_k^{(0)}$ 可以简单地根据对在核心编码器输出端解码的信号的短期谱分析进行估计。

还可以设想将这些初始化方法组合在一起。例如，可以通过将按照上式给出的 LPC 包络频谱相加、从根据 CELP 编码器编码的残差估计的短期频谱得出初始频谱 $\hat{X}_k^{(0)}$ 。

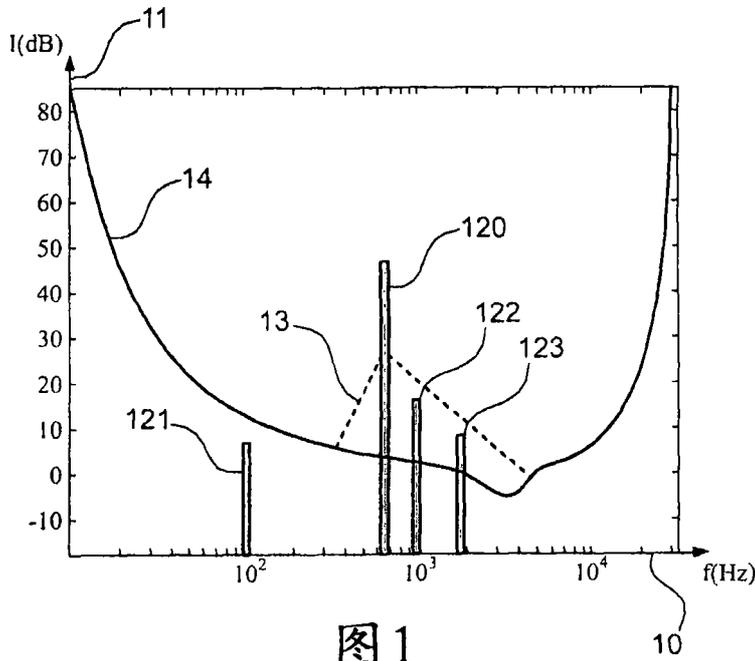


图 1

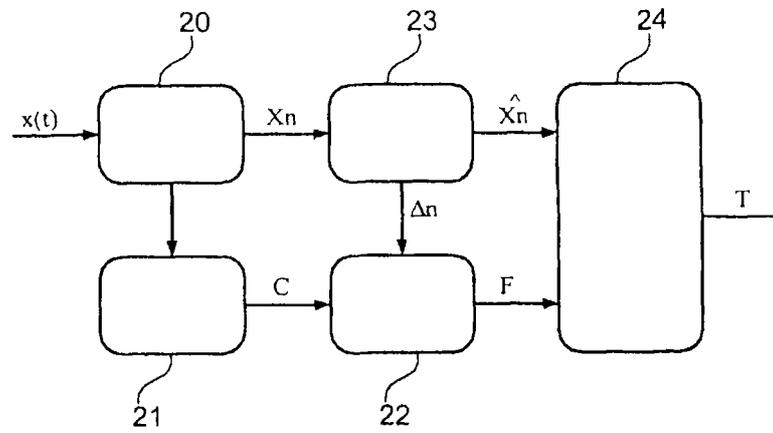


图 2

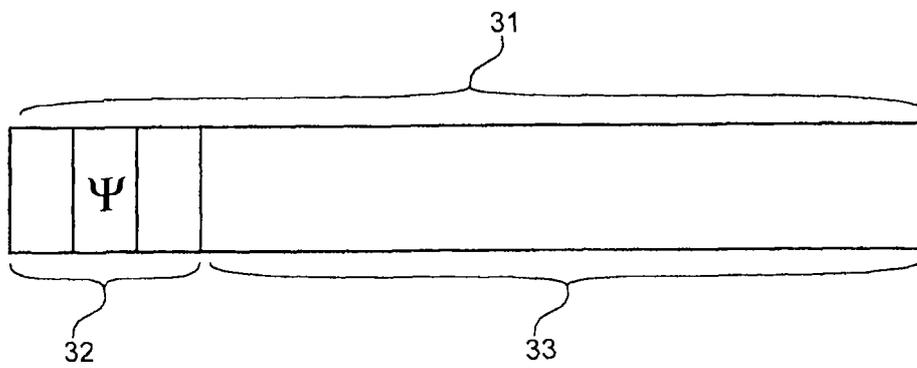


图 3

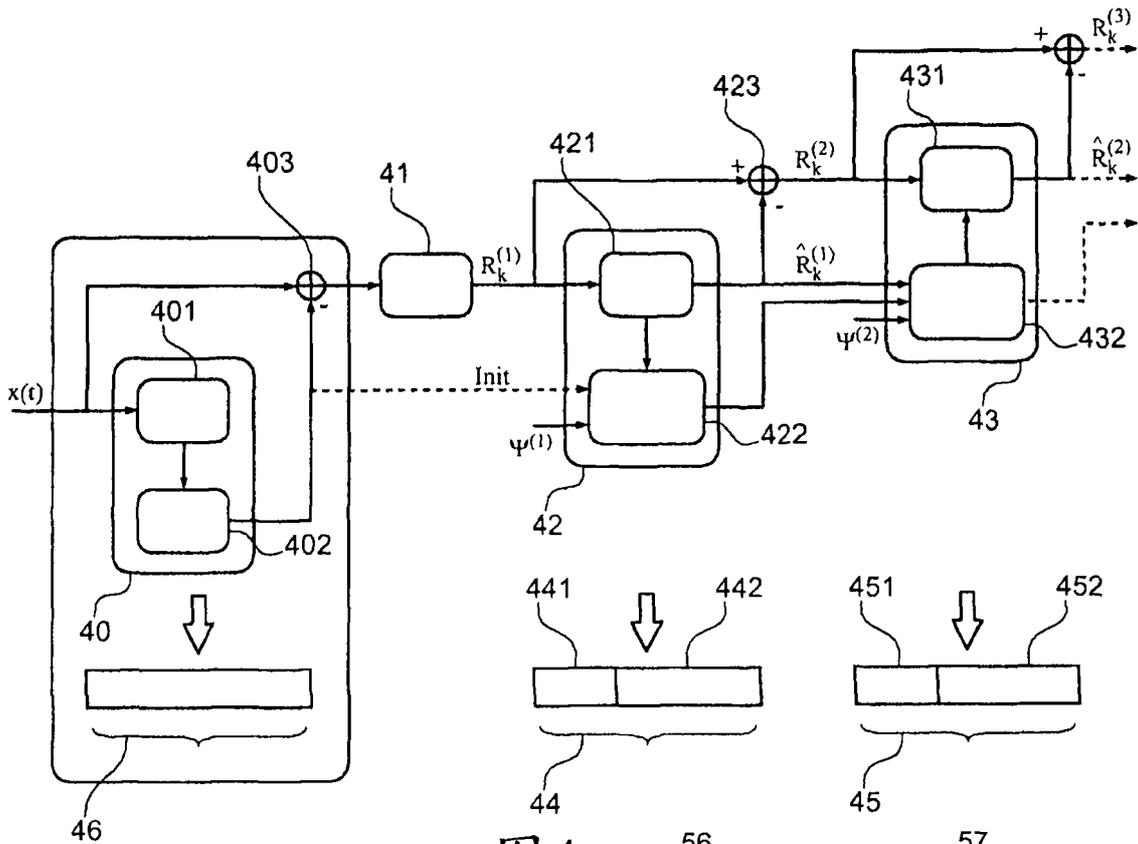


图 4

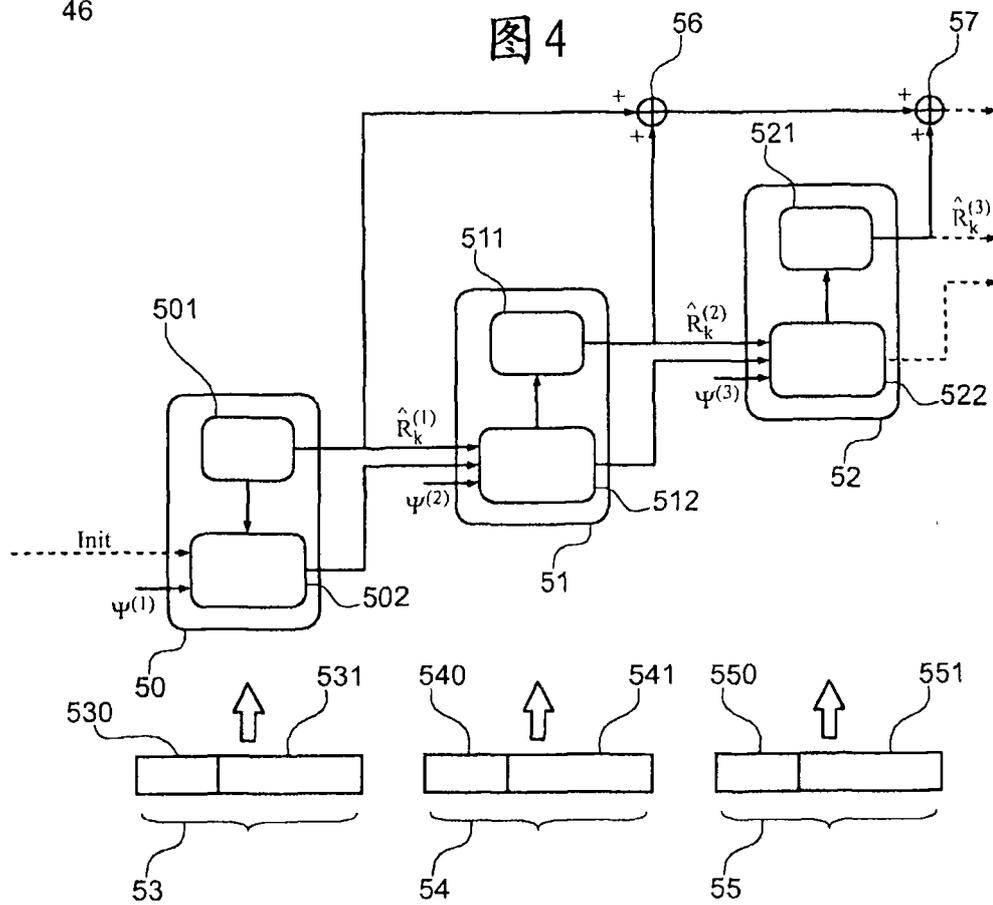


图 5

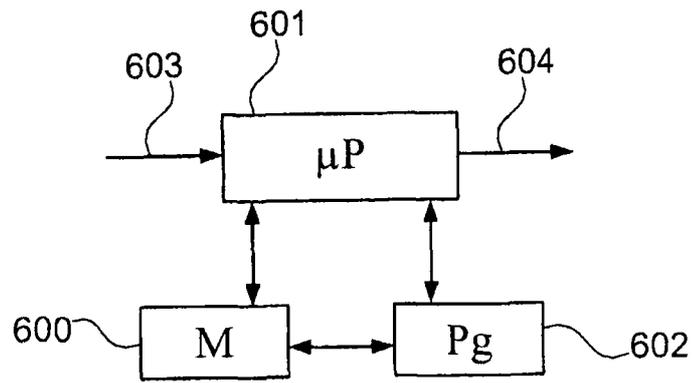


图 6A

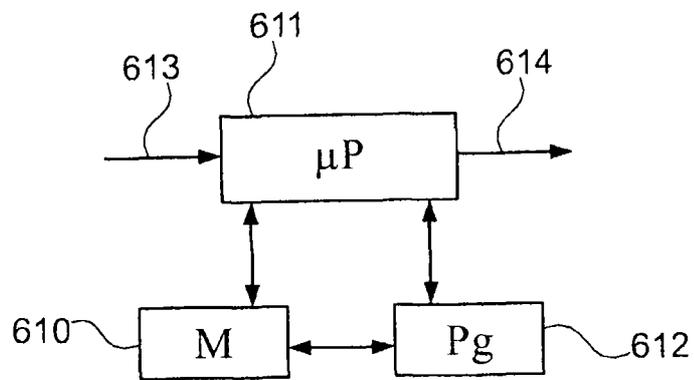


图 6B