



[12] 发明专利申请公开说明书

[21] 申请号 02828974.9

[43] 公开日 2005 年 6 月 8 日

[11] 公开号 CN 1625768A

[22] 申请日 2002.12.2 [21] 申请号 02828974.9
 [30] 优先权
 [32] 2002. 4. 18 [33] DE [31] 10217297. 8
 [86] 国际申请 PCT/EP2002/013623 2002. 12. 2
 [87] 国际公布 WO2003/088212 德 2003. 10. 23
 [85] 进入国家阶段日期 2004. 11. 16
 [71] 申请人 弗兰霍菲尔运输应用研究公司
 地址 德国慕尼黑
 [72] 发明人 拉尔夫·盖格 托马斯·思博尔
 卡尔海因兹·勃兰登堡
 朱尔根·赫尔 朱尔根·科洛尔
 乔吉姆·德格拉

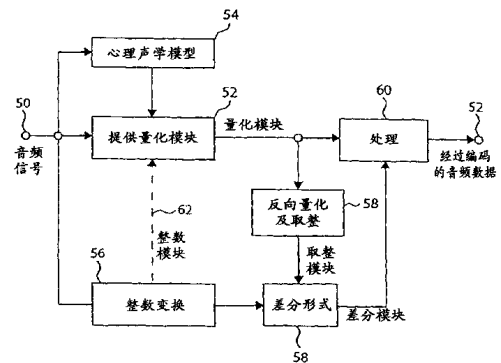
[74] 专利代理机构 中国国际贸易促进委员会专利商
 标事务所
 代理人 李 勇

权利要求书 7 页 说明书 23 页 附图 8 页

[54] 发明名称 对时间离散音频信号进行编码的装置和方法以及对已编码的音频数据进行解码的方法

[57] 摘要

时间离散音频信号被处理(52)，以提供一个带有量化频谱值的量化块。此外，使用整数变换算法(56)，从时间离散音频信号产生一种整数频谱表示。使用心理声学模型(54)产生的量化块被反向量化并取整(58)，以随后在整数频谱值和反向量化取整频谱值之间形成差值。在解码之后，这个量化块单独提供一种有损的心理声学编码/解码音频信号；而在解码中，这个量化块和结合模块一起提供一个无损或者几乎无损的编码和再次解码音频信号。通过在频域内产生差分信号，形成了一个简单的编码器/解码器结构。



1. 用于对时间离散的音频信号进行编码以得到编码后的音频数据的装置，包括：

用于使用心理声学模型（54）来提供被量化的时间离散的音频信号的量化块的装置（52）；

用于反向量化这个量化块，并对反向量化的频谱值取整，以获得被取整的反向量化的频谱值的取整块的装置（58）；

用于利用整数变换算法生成整数频谱值的整数块的装置（56），所述整数变换算法用于从整数时间离散采样模块来生成频谱值的整数块；

用于根据取整块和整数块之间频谱值的差形成差分块的结合装置（58），以获得具有差分频谱值的差分块；以及

用于处理量化块和差分块的装置（60），以产生包含量化块的信息和差分块的信息的编码的音频数据。

2. 如权利要求1所述的装置，其中用于提供的装置（52）通过一个 MDCT，从时间音频信号值的时间块产生一个 MDCT 频谱值的 MDCT 模块，并且

用心理声学模型来量化这个 MDCT 模块，以产生包含量化的 MDCT 频谱值的量化块。

3. 如权利要求2所述的装置，其中用于产生整数块的装置（56）在时间块上执行一个 IntMDCT，以产生包含 IntMDCT 频谱值的整数块。

4. 如前面任一权利要求所述的装置，其中用于提供的装置（52）用浮点转换算法计算量化块。

5. 如权利要求1至3中的任一项所述的装置，其中用于提供的装置（52）用通过用于生成的装置（56）产生的整数块来计算量化块。

6. 如前面任一权利要求所述的装置，

其中用于处理的装置（60）对量化块进行熵编码（60a），以获得熵

编码的量化块;

对取整块进行熵编码(60b), 以获得熵编码的取整块; 并且
将熵编码的量化块转换为表示编码音频数据的扩展数据流的第一扩展层, 并将熵编码取整块转换为扩展数据流的第二扩展层。

7. 如权利要求 6 所述的装置,

其中用于处理的装置(60)还根据量化的频谱值, 使用多个码表中的一个, 对量化块进行熵编码, 并且

其中用于处理的装置(60)还根据量化中可用的量化器的属性, 选择多个码表中的一个, 以产生用于对差分块进行熵编码的量化块。

8. 如前面任一权利要求所述的装置,

其中用于提供的装置(52)根据音频信号的属性, 选择多个窗口中的一个, 以对音频信号值的时间块进行窗口化; 并且

其中用于生成的装置(56)为整数转换算法进行相同的窗口选择。

9. 如权利要求 1 至 8 的任一项所述的装置,

其中用于生成的装置使用了一个整数转换算法, 包括:

用长度对应于 $2N$ 个时间离散的采样的窗口 (w) 对时间离散的采样进行窗口化, 以提供窗口化的时间离散采样, 通过能够从 N 个输入值产生 N 个输出值的变换, 将时间离散的采样变换为频谱表示, 其中窗口化过程包含下面的子步骤:

从窗口的四分之一部分选择 (16) 一个时间离散的采样, 并且从该窗口的另外四分之一部分选择一个时间离散的采样, 以得到时间离散采样的矢量;

应用一个旋转方阵, 其维数与矢量到矢量的维数相匹配, 其中旋转矩阵可以用多个提升矩阵来表示, 其中一个提升矩阵根据窗口 (w) 只包含一个元素, 并且不等于 1 或者 0, 其中应用于步骤包含下面的子步骤:

用提升矩阵与矢量相乘 (18), 得到第一个结果矢量;

用把实数映射为整数的取整函数 (r) 来取整第一个结果矢量的分量, 得到取整的第一个结果矢量; 并且

随后执行与另外一个提升矩阵相乘(22)并取整(24)的步骤,直到所有的提升矩阵都被处理完毕,得到一个旋转矢量,它包含来自窗口的四分之一一部分的整数窗口化采样和来自该窗口的另外四分之一一部分的整数窗口化采样,并且

执行对于窗口的剩余的四分之一一部分的所有时间离散采样进行窗口化的步骤,得到 $2N$ 个滤波后的整数值;并且

对于通过窗口的第二和第三个四分之一一部分的滤波后的整数采样值,通过整数 DCT,把 N 个窗口化的整数采样变换(14)为频谱表示,得到 N 个整数频谱值。

10. 如前面任一权利要求所述的装置,

其中提供量化块的装置(52)在量化步骤(52b)之前,用一个预测滤波器实现对于频率上频谱值的预测,以得到表示在量化之后的量化块的预测残留频谱值;

其中还提供了—个预测装置,它对整数块的整数频谱值在频率上进行预测,其中还提供了取整装置,以对由于表示取整块的整数频谱值得到的预测残留频谱值进行取整。

11. 如前面任一权利要求所述的装置,

其中时间离散音频信号包括至少两个信道:

其中用于提供的装置(52)用时间离散音频信号的频谱值来执行中心/边缘编码,以在中心/边缘频谱值的量化之后得到量化块,并且

其中用于生成整数块的装置(56)也执行对应于用于提供的装置(52)的中心/边缘编码的中心/边缘编码。

12. 如前面任一权利要求所述的装置,其中用于处理的装置(60)产生一个 MPEG-2 ACC 数据流,其中在一个区域中引入了用于整数变换算法的辅助数据辅助信息。

13. 如前面任一权利要求所述的装置,

其中用于处理的装置(60)输出经过编码的音频数据,作为带有多个扩展层的数据流。

14. 如权利要求 13 所述的装置,

其中用于处理的装置(60)在第一扩展层(81)中插入了关于量化块的信息,并且在第二扩展层(82)中插入了关于差分块的信息。

15. 如权利要求13所述的装置,

其中用于处理的装置(60)在第一扩展层中插入了关于量化块的信息,并且至少在第二和第三扩展层中插入了关于差分块的信息。

16. 如权利要求15所述的装置,

其中在第二扩展层中包含带有被减小的精度的差分频谱值,在高一級或者更高级可扩展层中包含差分频谱值的残留部分。

17. 如权利要求15或16所述的装置,

其中关于差分块的信息包含二进制编码差分频谱值;

其中用于差分频谱值的第二扩展层中包含从差分频谱值的最高有效位(MSB)到次高有效位(MSB-x)的多个比特;并且

其中在第三扩展层包含从次高有效位(MSB-x-1)到最低有效位(LSB)的多个比特。

18. 如权利要求17所述的装置,

其中时间离散音频信号用宽度为24比特的采样形式来表示,并且

其中用于处理的装置(60)在第二扩展层中插入差分频谱值的更高有效位的16个比特,在第三扩展层中插入差分频谱值的剩余的8个比特,这样解码器用第二扩展层达到了CD音质,其中如果采用第三扩展层,解码器就能达到演播室的音质。

19. 如权利要求15所述的装置,

其中用于处理的装置(60)在第二扩展层中插入了至少部分差分频谱值,表示低通滤波信号,在另外一个扩展层中插入了第二扩展层中的差分频谱值与初始差分频谱值之间的差值。

20. 如权利要求15或者19所述的装置,

其中用于处理的装置(60)在第二扩展层中插入了至少部分最高为某个截止频率的差分频谱值,并且在第三扩展层中插入了至少部分从某个截止频率到更高频率的差分频谱值。

21. 对时间离散的音频信号进行编码以得到编码音频数据的方法，包括：

使用心理声学模型(54)提供(52)一个量化的时间离散音频信号的频谱值的量化块；

反向量化(58)量化块，并且取整这个反向量化的频谱值，以得到取整反向量化频谱值的取整块；

使用一个整数变换算法来产生(56)一个整数频谱值的整数块，这个整数变换算法从整数时间离散采样块中产生频谱值的整数块；

根据在取整块和整数块之间的频谱差分值，形成(58)一个差分块，以得到带有差分频谱值的差分块；并且

处理(60)量化块和差分块，以产生包含关于量化块的信息和关于差分块的信息的编码音频数据。

22. 用于对已编码的音频数据进行解码的装置，该已编码的音频数据从一个时间离散音频信号产生，通过使用心理声学模型(54)提供(52)一个量化的时间离散音频信号的频谱值的量化块，通过反向量化(58)量化块并取整反向量化的频谱值，来获得取整后的反向量化频谱值的取整块，通过使用从整数时间离散采样的数据块来产生频谱值的整数块的整数变换算法，来产生(56)整数频谱值的整数块，通过根据在取整块和整数块之间的频谱值的差来形成(58)一个差分块，以得到差分频谱值的差分块，包括：

用于处理编码音频数据的装置(70)，得到一个量化块和差分块；

用于反向量化和取整这个量化块的装置(74)，以得到一个整数的反向量化的量化块；

用于以频谱值方式结合整数量化块和差分块的装置(78)，得到一个结合模块；

使用这个结合模块以及与整数变换算法相反的整数变换算法，来产生一个时间离散音频信号的时间表示的装置(82)。

23. 如权利要求 22 中所述的解码装置，

其中编码音频数据是可扩展的，并且包含多个扩展层；

其中用于处理这个编码音频数据的装置(70)从编码音频数据中确定出量化块, 作为第一扩展层, 并从编码音频数据中确定出差分块, 作为第二扩展层。

24. 如权利要求 22 所述的装置,

其中关于差分块的信息包含二进制的编码差分频谱值,

其中编码音频数据是可扩展的, 并且包含多个扩展层,

其中用于处理这个编码音频数据的装置(70)从编码音频数据中确定出量化块, 作为第一扩展层, 并且用降低了的精度提取出差分频谱值的表示, 作为第二扩展层。

25. 如权利要求 24 所述的装置,

其中用于处理这个编码音频数据的装置(70)提取出从最高有效位到次高有效位的多个比特作为第二扩展层, 其中次高有效位比一个差分频谱值中的最低有效位更高, 并且

用于生成离散音频信号的时间表示的装置(82)在使用整数变换算法之前, 以综合方式产生差分频谱值的缺失比特。

26. 如权利要求 25 所述的装置,

其中装置(82)为综合产生而执行第二扩展层的扩展, 其中在扩展中使用一个比例因子, 它等于 2^n , 其中 n 是不包含在第二扩展层中的次高有效位的数目, 或者为综合产生而使用抖动算法。

27. 如权利要求 22 所述的装置,

其中编码音频数据是可扩展的, 并且包含多个扩展层, 并且

用于处理这个编码音频数据的装置(70)从编码音频数据中确定出量化块, 作为第一扩展层, 并且确定低通滤波的差分频谱值, 作为第二个扩展层。

28. 如权利要求 22 或 27 所述的装置,

其中编码音频数据是可扩展的, 并且包含多个扩展层,

其中用于处理这个编码音频数据的装置(70)从编码音频数据中确定出量化块, 作为第一扩展层; 确定最高为第一个截止频率的差分频谱值, 作为第二扩展层, 其中第一个截止频率比可以在编码器中产生

的差分频谱值的最大频率要小。

29. 如权利要求 28 所述的装置，

其中用于生成时间表示的装置(82)把全长的整数变换算法的输入值设为预定值，这些值在第二扩展层的截止频率之上；并且在通过由对应于差分频谱值的最大频率和截至频率的比率选择的因子，使用反向整数变换算法之后，降低取样时间离散音频信号的时间表示，其中差分频谱值最大频率可由编码器产生。

30. 如权利要求 29 所述的装置，

其中所有在截止频率之上的输入值的预定值均为零。

31. 对已编码的音频数据进行解码的方法，其中已编码的音频数据通过提供、反向量化、生成、形成和处理，从时间离散的音频信号中产生，该方法包含：

处理(70)编码音频数据，以得到一个量化块和一个差分块；

反向量化(74)量化块并取整，以得到一个整数反向量化的量化块；

以频谱值的方式结合(78)该整数量化块和差分块，得到一个结合模块；以及

使用该结合模块，并使用与整数变换算法相反的整数变换算法，来产生(82)时间离散音频信号的时间表示。

32. 当程序在计算机上执行时，用于执行权利要求 21 的编码方法的具有程序代码的计算机程序。

33. 当程序在计算机上执行时，用于执行权利要求 31 的编码方法的具有程序代码的计算机程序。

对时间离散音频信号进行编码的装置和方法以及 对已编码的音频数据进行解码的方法

技术领域

本发明涉及音频编码/解码，尤其是涉及可扩展（scalable）的编码/解码算法，这种算法包含了一个心理声学的第一扩展层和一个包括用于无损解码的辅助音频数据的第二扩展层。

背景技术

现代音频编码方法，如 MPEG Layer3 (MP3) 或者 MPEG ACC，使用如所谓的修正离散余弦变换(MDCT)的变换来获得对音频信号的数据块式的频率表示。这样的音频编码器通常获得时间离散的音频采样的一个数据流。音频采样的数据流被窗口化（windowed）用以获取例如 1024 或者 2048 个窗口化的音频采样的窗口数据块。为了进行窗口化使用了多种窗口函数，例如正弦窗口等。

随后，窗口化的时间离散音频采样通过滤波器组被转换为频谱表示。原则上，傅立叶变换，或者用于特殊原因的多种傅立叶变换，如 FFT，或者前面阐述的 MDCT，都可以用于此。然后，在滤波器组输出端处的音频频谱值的数据块可以根据要求做进一步处理。在上面引用的音频编码器中，随后是音频频谱的量化，其中典型选择量化级，以使被量化引入的量化噪声在心理声学掩盖阈值之下，也就是说被“掩盖”住了。量化是一种有损编码。为了获得进一步的数据量缩减，量化的频谱值被熵编码，例如通过哈夫曼编码。通过添加辅助信息，如比例因子（scale factors）等，一个能够被存储或者传送的比特流通过比特流多路复用器从熵编码量化的频谱值中形成。

在音频解码器中，比特流被一个比特流分离多路复用器分割为编码量化的频谱值和辅助信息。熵编码的量化频谱值首先被熵解码，以

获得量化频谱值。经过量化的频谱值然后被反向量化，以获得包含量化噪声的解码频谱值，然而，这种量化噪声是在生理声学掩盖阈值之下的，因而是听不到的。然后这些频谱值通过合成滤波器组被转换为时间表示方式，以获得时间离散的解码音频采样。在合成滤波器组中，必须使用一种与变换算法相反的变换算法。而且，在频率-时间转换或者反变换后，窗口必须被取消。

为了获得良好的频率选择性，现代音频编码器典型地利用使用块重叠。这种情况在图 4a 中示出。首先，通过装置 402 取出例如 2048 个时间离散的音频采样，并窗口化。实现这种窗口的装置 402 具有 $2N$ 个采样的窗口长度，并在输出端提供了一个 $2N$ 个窗口化采样的数据块。为了获得窗口重叠，通过装置 404（仅仅是为了表述得更加清楚，该装置在图 4a 中与装置 402 被分开描述），形成了第二个 $2N$ 个窗口化采样的数据块。然而，被送入装置 404 的 2048 个采样不是紧接着第一个窗口的时间离散音频采样，而是包含了通过装置 402 窗口化了的采样的后半部分，此外仅包含了 1024 个“新”采样。在图 4a 中通过装置 406 示意性地说明了这个重叠，引起了 50% 的重叠度。然后，对通过装置 402 的 $2N$ 个窗口化采样输出和通过装置 404 的 $2N$ 个窗口化采样输出，分别用装置 408 和 410 实现 MDCT 算法。装置 408 根据已知的 MDCT 算法为第一个窗口提供了 N 个频谱值，而装置 410 也提供了 N 个频谱值，不过是用第二个窗口，其中第一个和第二个窗口之间有 50% 的重叠。

在解码器中，第一个窗口的 N 个频谱值，如图 4b 所示，被送入装置 412 来实现修正离散余弦反变换。同样的操作被作用于第二个窗口的 N 个频谱值。它们被送入装置 414，也实现了修正离散余弦反变换。装置 412 和装置 414 都分别为第一个窗口和第二个窗口提供了 $2N$ 个采样。

在装置 416 中，在图 4b 中以 TDAC（时域混迭取消）来表示，考虑到两个窗口是重叠的。特别地，第一个窗口的后半部分的一个采样 y_1 （也就是带有系数 $N+k$ ）与第二个窗口的前半部分的采样 y_2 （也

就是带有系数 k) 相加, 这样在输出端, 也就是解码器处生成 N 个已解码的时域采样。

需要注意的是, 通过也称为相加函数的装置 416 的功能, 在图 4a 所示的编码器中实现的窗口化在一定程度上被自动考虑, 所以在图 4b 所示的解码器中不必有明显的“反向窗口化”发生。

当通过装置 402 或者 404 实现的窗口函数被指定为 $w(k)$, 其中系数 k 代表时间系数, 必须满足的条件是平方后的窗口权重 $w(k)$ 与平方后的窗口权重 $w(N+k)$ 的和等于 1, 其中 k 的范围从 0 到 $N-1$ 。当使用正弦窗口时, 该窗口的权重遵循正弦函数的前半波, 这个条件始终满足, 因为任意角的正弦平方与余弦平方的和均为 1。

在图 4a 中描述的按照 MDCT 函数的窗口方法的缺点是, 通过将时间离散的采样相乘来窗口化, 当考虑它为一个正弦窗口的时候, 它由一个浮点数来达到, 因为一个在 0 到 180 度之间的角的正弦不会产生整数, 除非这个角等于 90 度。即便当整数时间离散采样被窗口化时, 在窗口化后也会产生浮点数。

因此, 即使当不使用心理声学编码时, 也就是当需要获得无损编码时, 为了进行适当的易于处理的熵编码, 在装置 408 或装置 410 的输出端处的量化也是必要的。

当已知的变换, 如在图 4a 基础上描述的那样, 被应用于无损音频编码, 需要使用非常好的量化, 以可以忽略由于浮点数取整而引起的结果误差, 或者误差信号需要例如在时域中被额外地编码。

现有技术中的概念, 也就是在其中量化被非常好地调整以使得由于浮点数取整而引起的结果错误可以被忽略, 例如在德国专利 DE 197 42 201 C1 中公开的那样。这里, 一个音频信号被转换为它的频谱表示并被量化, 以获得量化的频谱值。量化的频谱值然后被反向量化, 变换到时域, 并且被与原始的音频信号相比较。如果误差, 也就是原始音频信号与量化/反向量化后的音频信号之间的误差, 在一个误差阈值以上, 在反馈中量化器会被调整得更加精确, 然后再次进行比较。当低于误差阈值时, 停止迭代。可能仍然存在的残留信号被一个时域编

码器编码并被写入一个比特流，这个比特流除了时域编码的残留信号外还包括根据在迭代取消时候存在的量化器调整进行量化后的编码频谱值。需要注意的是，量化器不一定必须通过心理声学模型控制，以使编码的频谱值通常比由于采用心理声学模型而得到的频谱值量化得更为精确。

在出版物“A Design of Lossy and Lossless Scalable Audio Coding” (T. Moriya et al., Proc. ICASSP, 2000) 中描述了一个可扩展的编码器，这个编码器包括如一个 MPEG 编码器作为第一个有损数据压缩模块，此模块具有一个数据块形式的数字信号形式作为输入信号，并生成压缩的比特流。在另一个现有的本地解码器中编码再次被取消，并生成了一个编码/解码信号。这个信号通过从初始输入信号中减去编码/解码信号而与初始的输入信号相比较。误差信号然后被送到第二个模块，在那里使用了一个无损位转换器。这个转换有两步。第一步包括一个从二进制补码格式到符号数值格式的转换。第二步包括在一个处理块中从一个垂直数值序列到一个水平比特序列的转换。无损数据转换被执行以使零的数量最大化或者使一个序列中连续零的数量最大化，以便获得尽可能好的作为数字结果表示的时间误差信号。这一原理基于在出版物“Multi-Layer Bit Sliced Bit Rate Scalable Audio Coder” (103rd AES Convention, Preprint No. 4520,1997) 中阐明的比特片算法编码 (BSAC) 方案。

上述概念的缺点是用于无损扩展层的数据，也就是用于获得无损音频信号解码的辅助数据必须在时域中获得。这意味着获得为了获得时域的编码/解码信号需要包含频率/时间变换的完全解码，所以通过在原始音频输入信号与编码/解码音频信号之间的采样差异的形成来计算误差信号，编码/解码音频信号由于是心理声学编码因而是有损的。这个概念的缺点尤其在于在编码器生成音频数据流时，两种完全的时间/频率变换装置，如滤波器组或者如 MDCT 算法，都被要求用于前向的转换，另一方面，仅仅为了产生误差信号，需要一个完整的反向滤波器组或者一个完整的合成算法。因而，编码器除了它固有的

编码器功能，还必须具有完全的解码器功能。如果编码器是由软件实现的，则为此对存储性能和处理器性能都有所要求，从而导致编码器的实现增加了开销。

发明内容

本发明的目的在于提供一种花费较少的概念，利用这个概念，可以产生以一种以几乎无损的方式解码的音频数据流。

这个目标通过权利要求 1 中对时间离散的音频信号进行编码的装置，权利要求 21 中对时间离散的音频信号进行编码的方法，权利要求 22 中对已编码的音频数据进行解码的装置，权利要求 31 中对已编码的音频数据进行解码的方法，或者权利要求 32 或 33 中的计算机程序来实现。

本发明基于这样的发现，可以对音频信号进行无损解码的辅助音频信号可以通过如通常那样提供一个量化频谱值的数据块，然后对其进行反向量化来获得反向量化的频谱值来实现，反向量化的频谱值由于使用了心理声学模型量化因而是有损的。这些反向量化的频谱值然后被取整，以获得经过取整的反向量化的频谱值的取整块。作为形成差值的参考，按照本发明，使用了一种整数变换算法，此算法从一个整数时间离散采样块生成了只包含整数频谱值的频谱值整数块。按照本发明，现在在取整块和在整数块中的频谱值的结合是以频谱值的方式实现的，也就是说在频域内实现，所以在编码器本身不需要合成算法，也就是反向滤波器组或者反向 MDCT 算法等。由于整数变换算法和取整量化值，包含不同频谱值的结合块仅仅包含可以以某些已知方式熵编码的整数值。需要注意的是，任意的熵编码器都可以用于结合块的熵编码，如哈夫曼编码器和算法编码器等。

对量化块的量化频谱值编码也可以使用任意的编码器，如已知的现代音频编码器常用的工具。

值得注意的是，本发明的编码/解码概念与现代编码装置是兼容的，如窗口切换、TNS、或者多信道音频信号的中心/边缘编码。

在本发明的一个优选实施例中，用 MDCT 来提供一个使用心理声学模型量化的频谱值量化块。此外，最好使用一个所谓 IntMDCT 作为整数变换算法。

在本发明的替代实施例中，可以不使用通常的 MDCT，而 IntMDCT 可以作为 MDCT 的近似，即通过整数变换算法获得的整数频谱被用于心理声学量化器来获得量化的 IntMDCT 频谱值，此频谱值然后再次被反向量化并取整，以与原始的整数频谱值相比较。在这种情况下，只需要单一变换，也就是 IntMDCT 从整数时间离散采样产生整数频谱值。

典型地，处理器处理整数，或者每个浮点数被表示为整数。如果一个整数算法用于一个处理器，它可以无需对反向量化的频谱值取整，因为由于处理器取整值的算法，也就是在 LSB 精确度范围之内，即最低有效位，总是存在的。在这样的情况下，实现了完全的无损处理，也就是在被使用的处理器精度范围之内的处理。然而可选地，也可以取整到一个大致的精度，以使合成块中的差分信号被取整到一个由取整函数所确定的精确度。为了生成一个在数据压缩意义上几乎无损的编码器，在原本的处理系统取整外引入了取整，这样增强了灵活性，从而影响了编码无损的程度。

根据本发明的解码器本身在心理声学编码音频数据和辅助音频数据两方面特别突出，辅助音频数据从音频数据中抽取出，进行可能的熵解码，然后又做如下处理。首先解码器中量化块被反向量化，并且使用与编码器中一样的取整算法进行取整，这样随后可以被加到熵解码辅助音频数据上。在解码器中，然后心理声学压缩的音频信号的频谱表示和音频信号的无损表示同时存在，其中心理声学压缩的音频信号频谱表示被变换到时域，以获得一个无损的编码/解码音频信号，而所述无损表示通过使用与为获得无损，或者如上所述的那样，基本无损的编码/解码音频信号而使用的整数转换算法相反的整数转换算法变换到时域。

附图说明

本发明的上述及其他目标和特性将在下面与附图相结合的描述中更加清楚：

图 1 是用于处理时间离散的音频采样，以获得从中可确定整数频谱值的整数值的优选的装置的电路框图；

图 2 是一个在 Givens 旋转以及两个 DCT-IV 操作中的 MDCT 和反向 MDCT 的分解的示意图；

图 3 是在旋转和 DCT-TV 操作中有 50%重叠的 MDCT 分解的图例代表；

图 4a 是一个具有 MDCT 和 50%重叠的已知编码器的示意电路框图；

图 4b 是用于对图 4a 中生成的值进行解码的已知解码器的电路框图；

图 5 是一个优选的根据本发明的编码器的原理电路框图；

图 6 是一个可作为替代的优选的具有创造性的解码器的原理电路框图；

图 7 是一个具有创造性的优选解码器的原理电路框图；

图 8a 是具有一个第一扩展层和一个第二扩展层的比特流示意图；

图 8b 是具有一个第一扩展层和多个其它扩展层的比特流示意图；

图 9 是二进制编码差分频谱值的示意图，用于表示与差分频谱值的精确度（位）有关和/或与差分频谱值的频率（采样率）有关的可能扩展比率。

具体实施方式

在图 5 到 7 的基础上，下面将论及具有创造性的编码器电路（图 5 和图 6）或者一个具有创造性的优选的解码器电路（图 7）。图 5 所示的本发明的编码器包括一个输入端 50，时间离散的音频信号被送入这个输入端，还包括一个输出端 52，它输出已编码的音频数据。输入

端 50 处的时间离散的音频信号被馈入装置 52 以提供一个量化块，这个块在输出端提供了时间离散的音频信号的量化块，这个量化块包含使用生理声学模型 54 的时间离散频谱音频信号 50 的量化频谱值。本发明的编码器还包含使用一个整数变换算法 56 生成一个整数块的装置，其中这个整数算法对从整数时间离散采样生成整数频谱值是有效的。

具有创造性的编码器还包括用于从装置 52 对量化块输出进行反向量化的装置 58，并且，当需要和处理器精度不同的精度时，还包括一个取整函数。如同所述的一样，如果已经达到处理器系统的精度，则取整函数已经固有地包含在量化块的反向量化中，因为一个具有整数算法的处理器是无论如何不能够提供非整数值的。于是装置 58 提供了一种所谓的取整块，它包括固有地或者显式地被取整为整数的反向量化频谱值。取整块和整数块都被馈送到用于使用差异形成提供具有差分频谱值的差分的结合装置，在这里术语“差分块”意味着差分频谱值是包含整数块与取整块之间的差的数值。

从装置 52 输出的量化块以及从输出差异形成装置 58 的差分块都被送入处理装置 60，来实现如通常的量化块处理，并例如引起对差分块的熵编码。处理装置 60 在输出端 52 输出经过编码的音频数据，这些数据包括量化块的信息，还包括差分块的信息。

在第一个优选实施例中，如图 6 所示，时间离散的音频信号通过 MDCT 方法被转换为频谱表示，然后被量化。装置 52 用于提供量化块，具有 MDCT 装置 52a 和一个量化器 52b。

另外，最好用 IntMDCT56 作为整数转换算法来生成整数块。

在图 6 中，图 5 所示的处理装置 60 也作为比特流编码装置 60a 和熵编码器 60b 来描述，比特流编码装置 60a 是用于对装置 52b 输出的量化块进行比特流编码，熵编码器 60b 是用于对差分块进行熵编码。比特流编码器 60a 输出生理声学编码的音频数据，而熵编码器 60b 输出熵编码的差分块。模块 60a 和 60b 的两种输出数据块可以通过一种合适的方式结合为比特流，此比特流以生理声学编码的音频数据作为

第一扩展层，而把用于无损解码的辅助音频数据作为第二扩展层。这个经过扩展的比特流然后与图 5 所示的在编码器的输出端 52 处的已编码的音频数据相一致。

在一个替代的优选实施例中，可以不使用图 6 中的 MDCT 块 52a，因为它已在图 5 中通过虚线箭头 62 暗示了。在这种情况下，整数变换装置 56 提供的整数频谱被送到图 6 中形成差值的装置 58 和量化器 52b。由整数变换算法产生的频谱值在这里通过一种方式被用做通常的 MDCT 频谱的近似。这个实施例的好处在于，仅仅 IntMDCT 算法存在于编码器中，而不是 IntMDCT 和 MDCT 算法都需要存在。

再次参考图 6，需要注意的是，实框和实线代表遵循某一 MPEG 标准的一个普通音频编码器，而虚框和虚线则代表这样一个普通 MPEG 编码器的扩展。因此，可以看到不需要对普通 MPEG 编码器进行根本改变，而是通过增加整数变换器的方法来捕获无损编码的辅助音频数据，并不需要改变编码器/解码器的基本结构。

图 7 示出了一个用于对图 5 中输出端 52 处的已编码的音频数据输出进行解码的具有创造性的解码器的原理电路框图。它首先一方面分解为心理声学编码音频数据，另一方面分解为辅助音频数据。心理声学编码音频数据被送入一个普通的比特流解码器 70，而辅助音频数据，当在被编码器熵编码后，被编码器 72 熵编码。在图 7 中比特流解码器 70 的输出端处存在量化频谱值，这些频谱值原理上可以被送到与图 6 的装置中的反向量化器结构相同的反向量化器 74。如果需要达到一个与处理器精度不同的精度，在解码器中还提供了一个取整装置 76，取整装置 76 与图 6 的装置 58 一样，实现了将一个实数映射为一个整数的同样的算法或者同样的取整函数。在一个解码端结合器 78 中，经过取整的反向量化频谱值最好通过相加以频谱值的方式与熵编码辅助音频数据相结合，使得在解码器中，一方面反向量化频谱值出现在装置 74 的输出端处，另一方面整数频谱值出现在结合器 78 的输出端处。

然后，为了执行经过修正的离散余弦反变换，可以通过装置 80

把装置 74 的输出端处的频谱值变换到时域,以得到一个有损的心理声学编码和再解码的音频信号。为了执行反向的整数 MDCT (IntMDCT),可以通过装置 82 把合成器 78 的输出信号也变换到其时间形式,以产生一个无损的编码/解码音频信号,或者在采用一个更加粗略的取整的时候,能够产生一个几乎无损的编码和再解码的音频信号。

下面来看图 6 中的熵编码器 60b 一种特别优选的实施方式。在通常的现代 MPEG 编码器中,多个码表是根据量化频谱值的平均统计量来选择。最好在合成器 58 的输出端处的差分块使用相同的码表或者码书来进行熵编码。由于差分块的大小,即残留 IntMDCT 频谱,取决于量化的精度,因此熵编码器 60b 的码表选择可以在没有辅助边缘信息的情况下执行。

在一个 MPEG-2 AAC 解码器中,频谱系数,也就是量化频谱值,被分组为在量化块中的比例因子频带,其中频谱值以来自与比例因子频带相关的相应的比例因子的增益因子来加权。由于在这个已知的编码器概念中,一个非均匀的量化器被用于量化加权的频谱值,残留值的大小,也就是结合器 58 的输出端处的频谱值,不仅取决于比例因子,还取决于量化值自身。但是由于比例因子和量化频谱值都包含在由图 6 的装置 60a 生成的比特流中,也就是在心理声学编码音频数据中,最好根据差分频谱值的大小来实现解码器中的码书选择,以及在比特流中传输的比例因子和量化值的基础之上,确定出解码器中所使用的码表。由于在合成器 58 的输出端不需要传输辅助信息以对差分频谱值进行熵编码,熵编码仅仅导致数据率压缩,而不需要在数据流中扩展任何信号化比特作为熵编码器 60b 的辅助信息。

在一个遵循标准 MPEG-2 AAC 的音频编码器中,用窗口切换来避免瞬态音频信号域中的前向回波。这种技术基于在每半个 MDCT 窗口中分别选择窗口形状的可能性,能够在连续块中改变块的大小。同样的,IntMDCT 形式的整数变换算法(这种算法参照图 1 到 3 来解释)也在窗口化和在时域 MDCT 分解的混迭部分使用了不同的窗口形状

来执行。因而，为整数变换算法和生成量化块的变换算法最好使用相同的窗口判别。

在一个遵循 MPEG-2 AAC 的编码器中，也存在多种其它的编码工具，这里只介绍 TNS（时域噪声整形）和中间/边缘（CS）立体声编码。在 TNS 编码中，就在像 CS 编码中那样，在量化前对频谱值进行修正。接着，IntMDCT 值，也就是整数块，之间的差，以及量化 MDCT 值增加了。根据本发明，形成整数变换算法来接纳 TNS 编码和中间/边缘编码的整数频谱值。TNS 技术基于对 MDCT 值在频率上的自适应前向预测。通过一个信号自适应方式的普通 TNS 模块计算出的相同的预测滤波器最好也被用于预测整数频谱值，而如果其中产生了非整数值，则会使用向下取整，再次产生整数值。此取整最好发生在每个预测步骤之后。在解码器中，初始频谱可以通过使用反向滤波器和同样的取整函数再次重建。同样，CS 编码也可基于提升法通过使用具有角度 $\pi/4$ 的取整 Givens 旋转用于 IntMDCT 频谱值。因此，在解码器中的初始 IntMDCT 值是可以重建的。

需要注意的是，在以 IntMDCT 作为整数变换算法的优选实施例中，本发明的概念可以应用于一切基于 MDCT 的听觉适应性音频编码器。只是作为一个例子，这些编码器是根据 MPEG-4 AAC 可扩展性、MPEG-4 AAC 低时延、MPEG-4 BSAC、MPEG-4 Twin VQ、Dolby AC-3 等的编码器。

尤其需要注意的是，这个具有创造性的概念是反向兼容的。听觉适应性编码或解码器没有被改变，而仅仅是被扩展了。无损分量的辅助信息可以在以反向兼容方式的听觉适应性方式编码的比特流中传输，如在“辅助数据”域中的 MPEG-2 AAC。前面的听觉适应性解码器的附加部分在图 7 中以虚线表示，它可以与量化 MDCT 频谱和从听觉适应性解码器以无损方式获得的 IntMDCT 频谱一起来估计并重建辅助数据。

在无损或者几乎无损编码的补充下，心理声学编码的创造性的概念尤其适合产生、传输和解码可扩展数据流。已知可扩展数据流包含

许多不同的扩展层。其中，至少最低的扩展层可以被发送并与较高扩展层无关地进行解码。在数据的可扩展处理中，其它扩展层或者增强层被叠加到第一个扩展层或者基层上。一个完整的编码器可以产生可扩展的数据流，这个数据流具有第一可扩展层，原理上还有任意数目的其它可扩展层。可扩展性概念的一个优点是，假如有一个宽带传输信道可用，由编码器产生的可扩展数据流能够完全发送。也就是说，包括所有的可扩展层都可通过宽带传输信道来传输。但是，如果只有一个窄带的传输信道，经过编码的信号仍然可以通过传输信道发送，但是只能以第一扩展层或者某个数目的其它扩展层的形式来发送。其中其它扩展层的数目小于由编码器产生的所有扩展层数。当然，与信道连接并且适应信道的编码器可能已经产生基扩展层或第一扩展层以及多个与信道相关的其它可扩展层。

在解码器一端，可扩展概念也有一个优点，那就是反向兼容。这意味着只能处理第一扩展层的解码器忽略了数据流中的第二个以及其它扩展层，并且可以产生一个有用的输出信号。但是，如果解码器是一个典型的更加现代的解码器，能够处理扩展数据流中的多个扩展层，那么这个编码器能够作为基解码器来处理相同的数据流。

在本发明中，基本的可扩展性是量化的模块，即比特流编码器 60a 的输出，被写入到图 8 的第一个扩展层 81 中，当考虑图 6 的情况下，它包含心理声学编码的数据，例如帧。通过合成装置 58 产生的最好经过熵编码的差分频谱值被写入第二个扩展层中，这种简单的可扩展性在图 8a 中用 82 来表示。因此对帧来说，包含辅助音频数据。

如果从编码器到解码器的传输信道是宽带传输信道，扩展层 81 和 82 都可以发送到解码器。但如果这个传输信道是一个窄带传输信道，只有第一个扩展层是“符合”的，第二个扩展层可以在数据发送之前直接从数据流中移除，因此解码器只处理第一个扩展层。

在解码器一端，一个只能处理心理声学编码数据的“基解码器”可以在通过宽带信道收到第二个扩展层时直接忽略第二个扩展层。但如果这个解码器是一个含有心理声学解码算法和整数解码算法的完全的

解码器，那么它可以用第一个和第二个扩展层来解码，以产生无损编码和解码后的输出信号。

图 8a 中简要示出了本发明的一个优选实施例，用于帧的心理声学编码数据也被放在第一个扩展层中。图 8a 中的第二个扩展层被更精细地量化，使得从图 8 中的这个第二扩展层中出现多个扩展层，例如（更小的）第二扩展层、第三扩展层、第四扩展层等等。

从加法器 58 输出的差分频谱值尤其适合进一步的量化，如基于图 9 所示。图 9 简要示出了二进制编码的频谱值。图 9 中的每行 90 代表一个二进制编码的差分频谱值。在图 9 中差分频谱值根据频率来分类，在图上用箭头 91 来表示。一个差分频谱值 92 比差分频谱值 90 有更高的频率。图 9 中的表格中的第一列代表一个差分频谱值中的最高有效位；第二个数字代表有效位为 MSB-1 的比特；第三个数字代表有效位为 MSB-2 的比特。倒数第二列代表有效位为 LSB+2 的比特；倒数第一列代表有效位为 LSB+1 的比特；最后一列代表有效位数为 LSB 的比特，也就是一个差分频谱值的最低有效位。

在本发明的一个优选实施例中，差分频谱值的例如 16 个最高有效位在第二个扩展层中出现，以实现精确量化，这样如果希望的话，可以通过熵编码器 60b 进行熵编码。采用第二个扩展层的解码器在输出端以 16 比特的精度获得差分频谱值，这样第二扩展层和第一扩展层一起提供了一个 CD 音质的无损解码音频信号。已知存在 16 比特的 CD 音质音频采样。

另一方面，如果将演播室音质的音频信号提供给编码器，即，每个采样包含 24 比特的音频信号，则编码器可进一步产生包含差分频谱值的最后 8 比特的第三扩展层，并根据需要进行熵编码（图 6 的装置 60）。

一个完全的解码器获得第一扩展层、第二扩展层（差分频谱值 16 个最高有效位）和第三扩展层（差分频谱值 8 个次高有效位）的数据流，这个解码器可以提供一个无损的、演播室音质的编码/解码音频信号，也就是说，采用全部三个扩展层在解码器的输出端提供 24 比特的

采样字宽。

需要注意的是，演播室领域中音频信号比一般消费类领域音频信号有更长的采样字长。在消费类领域，音频 CD 中信号字宽是 16 比特，而在演播室领域中是 24 或 20 比特。

基于在 IntMDCT 领域缩放的概念，如前所述，所有三种精度（16 比特，20 比特或 24 比特）或者最小用 1 比特来量化的任意精度均可以被量化编码。

这里，用 24 比特精度表示的音频信号在借助于反向 IntMDCT 在整数频域中表示，并且和听力适应的基于 MDCT 的音频编码输出信号量化结合。

用于无损表示的整数差分值现在不是在一个扩展层中完全编码，而是首先以一种比较低的精度来编码。仅在一个其它扩展层中发送为精确的表达所需的残留值。然而一种替代方案是，一个差分频谱值可以被完整的表示，即在其它扩展层中例如用 24 比特来表示，这样对于解码这个其它的可扩展层，则不再需要下面的扩展层。然而，这种情况会导致更高的比特流大小，但是当传输信道的带宽不存在问题时，在解码器端就会简化，因为在解码器中可扩展层不再需要结合起来，对解码始终采用一个扩展层就足够了。

例如如果低 8 位 LSB，如图 9 所示，在开始时不再发送，就能实现在 24 比特和 16 比特之间的可扩展性。

为了将用较低精度所传输的值反变换到时域，被传输的值最好被扩展回初始区域，例如 24 比特，例如用 2^8 乘以所传输的值。一个反向的 IntMDCT 被应用到对应的扩展回的值。

在根据本发明的频域中的精度量化中，还最好利用 LSB 中的冗余。例如如果一个音频信号在上部频域有很小的能量，这在 IntMDCT 频谱中用很小的值来表示，例如这些值大大小于可以例如用 8 比特表示的值（-128,, 127），在 IntMDCT 频谱的 LSB 值的可压缩性中也体现了这种情况。而且，需要注意的是：在很小的差分频谱值中，从 MSB 到 MSB-1 的多个比特典型地均等于零；在有效位为 MSB-n-1

的比特之前，二进制编码的差分频谱值中的第一个 1 并不存在。这种情况下，当在第二个可扩展层中的差分频谱值只包含零的时候，熵编码尤其适合进一步的数据压缩。

按照本发明的另一个实施例，对于图 8a 的第二扩展层 82 最好使用采样率扩展性。采样率扩展性通过最大为包含在第二扩展层中的第一截止频率的差分频谱值来实现，如图 9 右边所示，而在其它扩展层中，包含频率位于第一截止频率和最大频率之间的差分频谱值。当然，可以实现进一步的扩展，以在整个频域形成多个扩展层。

在本发明的一个优选实施例中，图 9 中的第二个扩展层包括频率最大为 24 kHz 的差分频谱值，对应于 48 kHz 的采样率。第三扩展层包括从 24 kHz 到 48 kHz 的差分频谱值，对应于 96 kHz 的采样率。

需要进一步注意的是，在第二扩展层和第三扩展层中，不是一个差分频谱值中的所有位都需要编码。在合成扩展性的其它形式中，第二扩展层可包含最大为某一截止频率的差分频谱值的从 MSB 到 MSB-X 的位。第三扩展层然后可以包含从第一截止频率到最高频率的差分频谱值的从 MSB 到 MSB-X 的位。第四扩展层可包含最大为截止频率的差分频谱值的剩余位。最后一个扩展层包含较高频率的差分频谱值的剩余位。这个概念将会使图 9 中的表格被分为四个象限，每个象限代表一个扩展层。

在频率的可扩展性中，在本发明的一个优选实施例中，描述了一个位于 48 kHz 和 96 kHz 采样率之间的可扩展性。96 kHz 的采样信号首先只在无损扩展层的 IntMDCT 区域中编码一半，并被传输。如果上半部分不被另外传输，在解码器中它被假定为零。在反向 IntMDCT 中（与编码器同样长度），产生了一个 96kHz 的信号，这个信号在上面的频域不包含能量，因而可能在没有质量损失的情况下以 48 kHz 被二次采样。

考虑到可扩展层的大小，图 9 具有固定边界的象限中差分频谱值最好在上面量化，因为在一个扩展层中，实际上只需要包含例如 16 位或者 8 位或者最大为截止频率或高于截止频率的频谱值。

一种作为替换的比例在某种程度上“软化”了图 9 的象限边界。在频率可扩展性的例子中，这意味着不因为在截止频率前的差分频谱值没有改变并且在截止频率后为零，就应用所谓的“砖墙低通”。相反的，差分频谱值也可通过已经有些阻碍低于截止频率的频谱值的任意低通来滤波，但是，在截止频率以上，差分频谱值还仍然有能量，虽然能量在降低。在由此生成的扩展层中，还包含在截止频率以上的频谱值。然而，由于这些频谱值相对较小，它们可以被有效地进行熵编码。在这种情况下最高扩展层具有在完全差分频谱值和包含在第二扩展层的频谱值之间的差。

精确量化在某种程度上也可以同样被软化。第一扩展层也包含例如多于 16 位的频谱值，其中在下一个扩展层中仍然具有这个差别。通常来讲，第二扩展层具有精度更低的差分频谱值，而在下一个扩展层中，其余的，也就是完全频谱值和第二可扩展层中包含的频谱值之间的差被传输。通过这种方法，实现了可变精度缩减。

具有创造性的编码或解码方法更适于存储在具有电子可读性控制信号的电子存储媒体中，如软盘，其中控制信号可以与一个可编程的计算机系统配合，从而执行编码和/或解码方法。换句话说，当程序产品在计算机上执行时，存在一个具有存储在机器可读载体的计算机代码的计算机程序产品，以实现编码和/或解码方法。当程序在计算机中执行时，本发明的方法可以通过具有执行本发明的方法的计算机代码的计算机程序来实现。

下面，作为一个整数变换算法的例子，需要介绍在“**Audio Coding Based on Interger Transforms**” (111th AES convention, New York, 2001) 中描述的 IntMDCT 变换算法。由于 IntMDCT 有 MDCT 算法的吸引人的特性，如音频信号的良好频谱表示、严格的取样和块重叠，IntMDCT 尤其受到青睐。一种通过 IntMDCT 对 MDCT 的良好近似可以仅仅使用在图 5 的编码器中的一个变换算法，如图 5 的箭头 62 所示。在图 1 到 4 的基础之上解释了这种特别形式的整数变换算法的重要属性。

图 1 示出了为处理表示音频信号的时间离散的采样的具有创造性的优选的装置，以获得使 IntMDCT 整数变换算法有效的整数值。时间离散的采样被窗口化并且可选地被图 1 所示的装置转换成频谱表示。被送入装置的输入端 10 的时间离散的采样被一个长度为 $2N$ 时间离散采样的窗口 w 窗口化，以在输出端 12 获取整数窗口化采样，这些采样适合于通过变换装置、尤其是用于执行整数 DCT 的装置 14 转化为频谱表示。整数 DCT 用于从 N 个输入值产生 N 个输出值，这与图 4a 的 MDCT 函数 408 相反，函数 408 根据 MDCT 等式从 $2N$ 个窗口化值只产生 N 个频谱值。

为了窗口化时间离散采样，首先在装置 16 中选择两个时间离散的采样，它们一起代表一个时间离散采样的矢量。装置 16 选择的一个时间离散采样位于窗口的第一象限。另一个时间离散采样位于窗口的第二象限，在图 3 的基础上它被解释得更加详细。对于装置 16 生成的矢量，应用一个 2×2 维的矩阵旋转，其中这个操作不是立即执行的，而是通过多个所谓的“提升矩阵”来执行。

一个提升矩阵具有只包含一个与窗口 w 有关的元素和不等 0 或者 1 的属性。

在“Factoring Wavelet Transforms Into Lifting Steps” (Ingrid Daubechies 和 Wim Sweldens, preprint, Bell Laboratories, Lucent Technologies, 1996) 中描述了由小波变换到提升步骤的因式分解。总体来讲，一个提升方案是具有同样低通或者高通滤波器的完美重建滤波器对之间的简单关系。每对互补滤波器都可以被因式分解为提升步骤。这对于 Givens 旋转尤其适用。考虑多相矩阵是 Givens 旋转的情形。然后，应用下面的公式：

$$\begin{pmatrix} \cos a & -\sin a \\ \sin a & \cos a \end{pmatrix} = \begin{pmatrix} 1 & \frac{(\cos a - 1)}{\sin a} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \sin a & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{(\cos a - 1)}{\sin a} \\ 0 & 1 \end{pmatrix} \quad (1)$$

等号右边的三个提升矩阵每个都有 1 作为主对角线元素。此外，在每个提升矩阵中，不在主对角线上的元素等于 0，不在主对角线上的元素与旋转角 a 有关。

现在向量与第三个提升矩阵相乘，也就是乘以上式中最右边的提升矩阵，得到第一个结果向量，在图 1 中用装置 18 来描述这个过程。如图 1 中通过装置 20 所示，用一个任意的取整函数对第一个结果向量取整，这个取整函数将一组实数映射为一组整数。在装置 20 的输出端处得到了取整后的第一个结果向量。这个取整后的第一个结果向量被送到装置 22，与中间的一项相乘，也就是乘以右边第二项，得到第二个结果向量，然后再用装置 24 取整得到取整后的第二个结果向量。取整后的第二个结果向量送至装置 26 与上述等式最左边的提升矩阵相乘，也就是第一项，来得到第三个结果向量，最后依然用装置 28 取整，最后在输出端 12 处得到整数窗口化采样，如果希望得到其频谱表示，则需要通过装置 14 对其进行处理，从而在频谱输出端 30 处得到整数频谱值。

装置 14 最好作为整数 DCT 来实现。

根据长度为 N 的类型 4 (DCT-IV)，离散余弦变换用下式给出：

$$X_i(m) = \sqrt{\frac{2}{N}} \sum_{k=0}^{N-1} x(k) \cos\left(\frac{\pi}{4N}(2k+1)(2m+1)\right) \quad (2)$$

DCT-IV 的系数形成一个标准正交的 $N \times N$ 矩阵，如出版物“Multirate System And Filter Banks” (P.P.Vaidyanathan, Prentice Hall, Englewood Cliffs, 1993) 中所述，每一个正交 $N \times N$ 矩阵可以分解成 $N(N-1)/2$ 个 Givens 旋转。需要注意的是，也可以进一步分解。

对于不同 DCT 算法的分类，可以参考 H.S.Malvar 的“Signal Processing With Lapped Transforms”一书，1992 年 Artech House 出版社出版。一般来说，DCT 算法根据它们的基函数类型来区分。而在这里优选的 DCT-IV 中包含非对称的基函数，也就是说，一个 $1/4$ 余弦波，一个 $3/4$ 余弦波，一个 $5/4$ 余弦波，一个 $7/4$ 余弦波等等，这种离散余弦变换，例如类型 II (DCT-II)，具有轴对称和点对称的基函数。零级基函数是一个直流分量，第一级基函数是半个余弦波，第二级基函数是整个余弦波，等等。由于在 DCT-II 中特别考虑直流分量，它应用在视频编码中而不是用在音频编码中，因为与视频编码不同的是，音频编码中的直流分量是不相关的。

下面来解释 Givens 旋转的旋转角 α 如何与窗口函数有关。

窗口长度为 $2N$ 的一个 MDCT 可以减至长度为 N 的 IV 型离散余弦变换。这可以通过在时域内执行 TDAC 操作，然后应用 DCT-IV 来实现。由于 50%重叠，用于块 t 的左半部窗口和先前的块，也就是块 $t-1$ 的右半部重叠。两个连续块 t 和 $t-1$ 的重叠部分在时域中，即在转换之前，也就是在图 1 的输入 10 和输出 12 之间，进行预处理，如下：

$$\begin{pmatrix} \tilde{x}_t(k) \\ \tilde{x}_{t-1}(N-1-k) \end{pmatrix} = \begin{pmatrix} w(\frac{N}{2}+k) & -w(\frac{N}{2}-1-k) \\ w(\frac{N}{2}-1-k) & w(\frac{N}{2}+k) \end{pmatrix} \begin{pmatrix} x_t(\frac{N}{2}+k) \\ x_t(\frac{N}{2}-1-k) \end{pmatrix} \quad (3)$$

字母上面标有波浪线的数值是图 1 的输出端 12 处的值，上式中没有标有波浪线的 x 值代表输入端 10 处的值或者装置 16 后面的用于选择的值。系数 k 的取值范围从 0 到 $(N/2)-1$ ， w 代表窗口函数。

从窗口函数 w 的 TDAC 条件可知有下面关系：

$$w\left(\frac{N}{2}+k\right)^2 + w\left(\frac{N}{2}-1-k\right)^2 = 1 \quad (4)$$

对于某些角度 α_k ， $k=0, 1, \dots, (N/2)-1$ ，这个在时域内的预处理可以写成 Givens 旋转，这在前面已经解释了。

Givens 旋转的角度 α 与窗口函数 w 的关系如下：

$$\alpha = \arctan[w(N/2-1-k)/w(N/2+k)] \quad (5)$$

需要注意的是，只要符合 TDAC 条件，任意的窗口函数 w 都可以应用。

下面，以图 2 为基础，描述了一个级联的编码器和解码器。通过一个窗口一起“窗口化”的时间离散采样 $x(0)$ 到 $x(2N-1)$ 首先被图 1 中的装置 16 来选择，使得采样 $x(0)$ 和 $x(N-1)$ ，即来自窗口的第一个四分之一部分的采样和来自窗口的第二个四分之一部分的采样被选择，以在装置 16 的输出端处形成矢量。交叉的箭头表示对装置 18, 20 或 22, 24 或 26, 28 提升相乘和相继取整，以在 DCT-IV 块的输入端得到整数窗口化的采样。

如上所描述，当第一个矢量被处理的时候，第二个矢量也从采样 $x(N/2-1)$ 和 $x(N/2)$ 中选中，也就是说，又一个来自窗口的第一个四分之一

一部分的采样和来自窗口的第二个四分之一一部分的采样，再一次通过图 1 中所描述的算法处理。所有其他的来自于窗口第一个四分之一一部分和第二个四分之一一部分的采样对均被类似处理。第一个窗口的第三个和第四个四分之一一部分被同样地处理。如图 2 所示，在输出端 12 处具有 N 个“窗口化”的整数采样，它被送至 DCT-IV 变换。特别的，第二个和第三个四分之一一部分的“窗口化”整数采样被送至 DCT。窗口的第一个四分之一一部分的“窗口化”整数采样与前一个窗口的第四个四分之一一部分的“窗口化”整数采样一起被送入前面的 DCT-IV 中进行处理。类似的，图 2 中第四个四分之一一部分的“窗口化”整数采样与后一个窗口的第一个四分之一一部分的“窗口化”整数采样一起被送至 DCT-IV 变换。图 2 中所示的中央整数 DCT-IV 变换 32 提供了 N 个整数的频谱值 $y(0)$ 到 $y(N-1)$ 。由于窗口化过程和变换过程提供了整数的输出值，因此不需要反向量化就可以将这些整数频谱值直接进行熵编码。

在图 2 的右半边描述了一个解码器。这个解码器包含反向变换和“反向窗口化”，它以与编码器相反的方式工作。已知对于 DCT-IV 的反向变换来说，需要使用到如图 2 所示的反向 DCT-IV。如图 2 所示，为了再一次在装置 34 的输出端或者前一次和下一次变换中从整数“窗口化”采样中产生时间离散音频采样 $x(0)$ 到 $x(2N-1)$ ，用前一次和后一次的变换的值对解码器 DCT-IV34 的输出值进行反向处理。

输出端的操作通过一个反向 Givens 旋转来完成，即块 26, 28 或者 22, 24 或者 18, 20 是在一个相反的方向通过。基于等式 1 的第二个提升矩阵可以描述得更加详细。当（在编码器中）第二个结果矢量通过将取整后的第一个结果矢量与第二个提升矩阵相乘（装置 22）而形成的时候，有以下的结果：

$$(x, y) \mapsto (x, y + x \sin a) \quad (6)$$

等式 6 右边的值 x, y 是整数。然而这不适用于值 $x \sin a$ 。这里，需要介绍一下取整函数 r ，它以如下的等式表示：

$$(x, y) \mapsto (x, y + r(x \sin a)) \quad (7)$$

这个操作执行了装置 24 的功能。

解码器中的反向映射可以定义如下：

$$(x', y') \mapsto (x', y' - r(x' \sin a)) \quad (8)$$

由于在取整操作之前的减号，很明显提升步骤的整数近似可以被反向，而不会引入错误。对这三个提升步骤中任何一个的近似的应用都导致了 Givens 旋转的整数近似。（编码器中的）取整旋转可以（在解码器中）被反向，而不会引入错误，即反向取整顺提升步骤以相反的顺序通过，也就是说，图 1 的算法在解码的时候是自下向上执行的。

如果取整函数 r 是点对称的，反向取整的旋转与角 $-\alpha$ 的取整旋转是相同的，如下：

$$\begin{pmatrix} \cos a & \sin a \\ -\sin a & \cos a \end{pmatrix} \quad (9)$$

用于解码器的提升矩阵，即用于反向 Givens 旋转，在这种情况下可由等式 (1) 直接得到，仅需简单地将“ $\sin a$ ”项替换为“ $-\sin a$ ”。

在下面，在图 3 的基础之上，再次提到具有重叠窗口 40 到 60 的普通 MDCT 的分解。窗口 40 到 60 分别重叠 50%。每个窗口，首先窗口的第一和第二个四分之一部分内、或者在窗口的第三和第四个四分之一部分内的 Givens 旋转被执行，如箭头 48 所示。然后，被旋转的值，也就是窗口化的整数采样，被送入一个 N 到 N 的 DCT，使得一个窗口的第二和第三个四分之一部分或者下一个窗口的第四和第一个四分之一部分一起通过 DCT-IV 算法转换为频谱表示。

所以，通常的 Givens 旋转被分解为提升矩阵，这些矩阵被顺序执行，其中在每次提升矩阵相乘之后引入一个取整的步骤，使得浮点数在它们产生后就立即被取整，这样在每次结果矢量与提升矩阵相乘之前，结果矢量只有整数。

输出值总是整数，最好也使用整数输入值。这不代表对本发明的局限，因为每个作为示例的 PCM 采样，由于它们存储在一张 CD 上，是整数值，其取值范围是根据位的宽度变化的，也就是说，根据时间离散数字输入值是十六位还是二十一位来变化。然而，如所阐述的一样，通过以相反的顺序执行反向旋转，整个过程是可以反向进行的。因此，存在一个具有完美重建的 MDCT 整数近似值，即无损转换。

所示转换提供了整数输出值而不是浮点值。它提供了一个完美的重建，所以当先执行一个前向转换、然后执行一个后向转换的时候，没有引入错误。这个转换，按照本发明的一个优选实施例，是对修正离散余弦变换的替换。然而，其他转换方法也可以通过整数的方式执行，只要分解为旋转和将旋转分解为提升步骤是可能的。

整数 MDCT 有 MDCT 的大部分优良特性。它有一个重叠的结构，由此可得到比在无重叠块转换中更好的频率选择性。由于 TDAC 函数，转换前的窗口化已经考虑了这个函数，维持了严格的采样，使得代表一个音频信号的所有频谱值等于输入采样的总数。

与一个普通的提供浮点采样的 MDCT 相比，在描述的优选的整数变换中，仅在具有很小的信号强度的频谱区域中，与普通 MDCT 相比，噪声增强了，而这个噪声增强的并没有使它自己成为一个重要的信号强度。为此，整数处理有助于有效的硬件实现，因为只使用了乘法步骤，而乘法可以很容易地分解为移位和加法步骤，这两种操作在硬件中都是很容易快速实现的。当然，软件实现也是可行的。

整数变换提供了音频信号的一个良好的频谱表示，并且仍然保留在整数区域。当它被应用于一个音频信号的语音部分时，会导致良好的能量聚集。通过这种方法，一个有效的无损编码方案可以通过用如图 1 所示简单的级联窗口化/转换来实现。尤其，使用溢出值的堆栈编码是很受欢迎的，如在 MPEG AAC 中使用的一样。最好通过使用二的特定次方来缩减所有的值直到它们满足一个所希望的码表，然后对忽略的最低有效位进行编码。与使用更大的码表的替代方法相比，考虑到存储码表所需要的存储消耗，这个方法更好。也可以通过只简单地省略某些最低有效位获得一种几乎无损的编码器。

尤其对于语音信号，整数频谱值的熵编码使高编码增益成为可能。对于信号的瞬态部分，编码增益很低，即由于瞬态信号的平坦频谱，也就是说，由于一小部分等于或几乎等于 0 的频谱值。如在 J. Herre, J.D. Johnston 的“Enhancing the Performance of Perceptual Audio Coders by Using Temporal Noise Shaping (TNS)”^{101st} AES

Convention, Los Angeles, 1996, preprint 4384 中所描述, 然而这种平坦性可能通过用频域内的线性预测而被利用。有一个替代方案是用开环预测, 还有一个替代方案是用闭环预测。第一种方案, 即开环预测器, 被称为 TNS。预测后的量化导致结果量化噪声适应于音频信号的时域结构, 因此阻止了在心理声学音频编码器中的前向回波。对于无损音频编码, 第二种方案更适合, 也就是闭环预测器, 因为闭环预测允许输入信号的精确重建。当这一技术被应用于所生成的频谱时, 在预测滤波器的每级后必须执行一个取整步骤, 以使之保留在整数区域内。通过使用反向滤波器和同样的取整函数, 初始的频谱可以精确地产生。

为了利用数据缩减中的两条信道之间的冗余, 当使用一个 $\alpha/4$ 角度的取整旋转时候, 在无损方式中也可以使用中间-边缘编码。与计算立体声信号左右声道之间的总数和差的方法相比较, 这个取整旋转的好处是能够维持能量。使用所谓的结合立体声编码的技术可以为每个波段被打开或者关闭, 如同在标准 MPEG AAC 中也是这样实现的。为了能够更加灵活地减小两个信道之间的冗余, 还可考虑其它旋转角度。

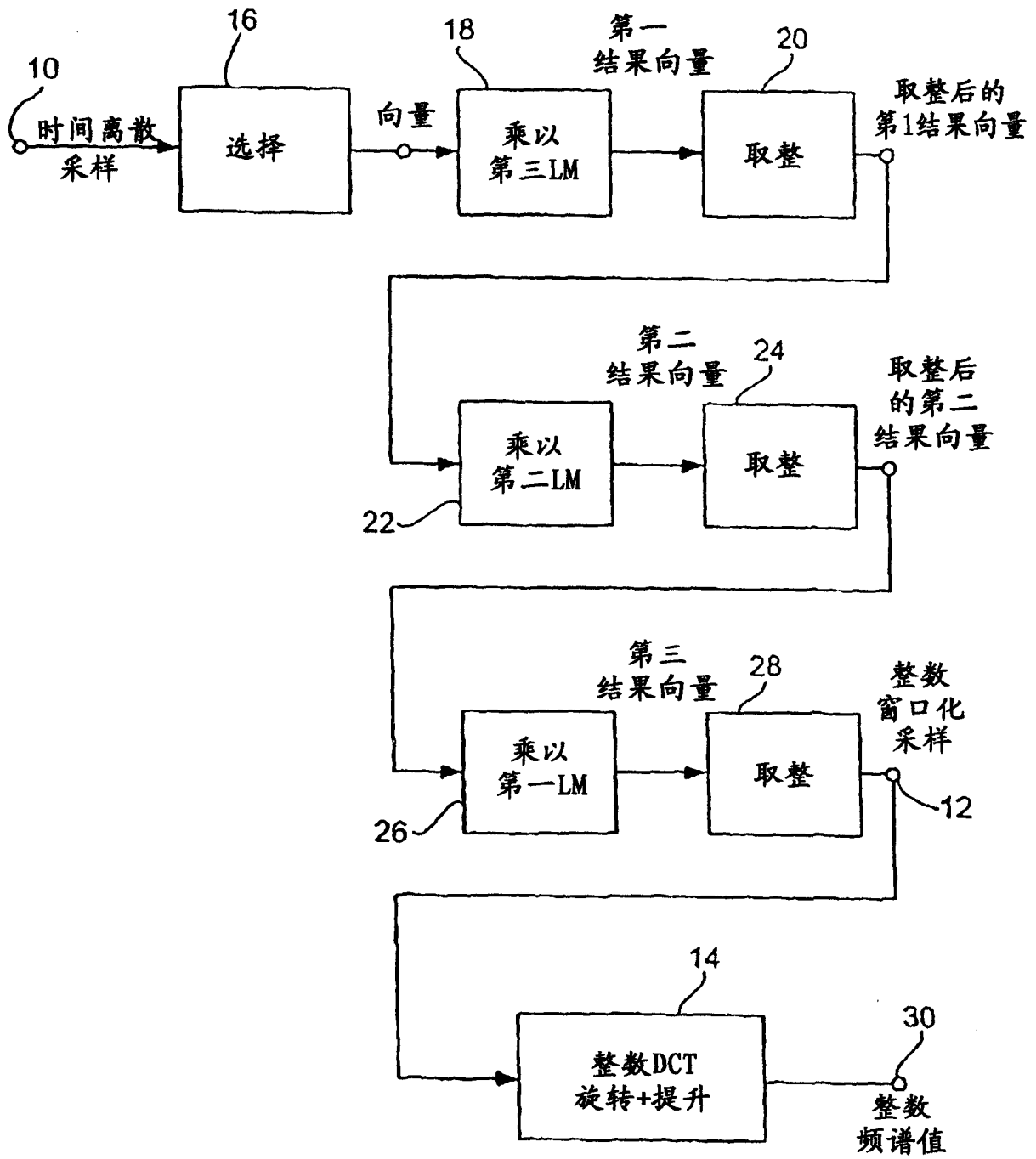


图1

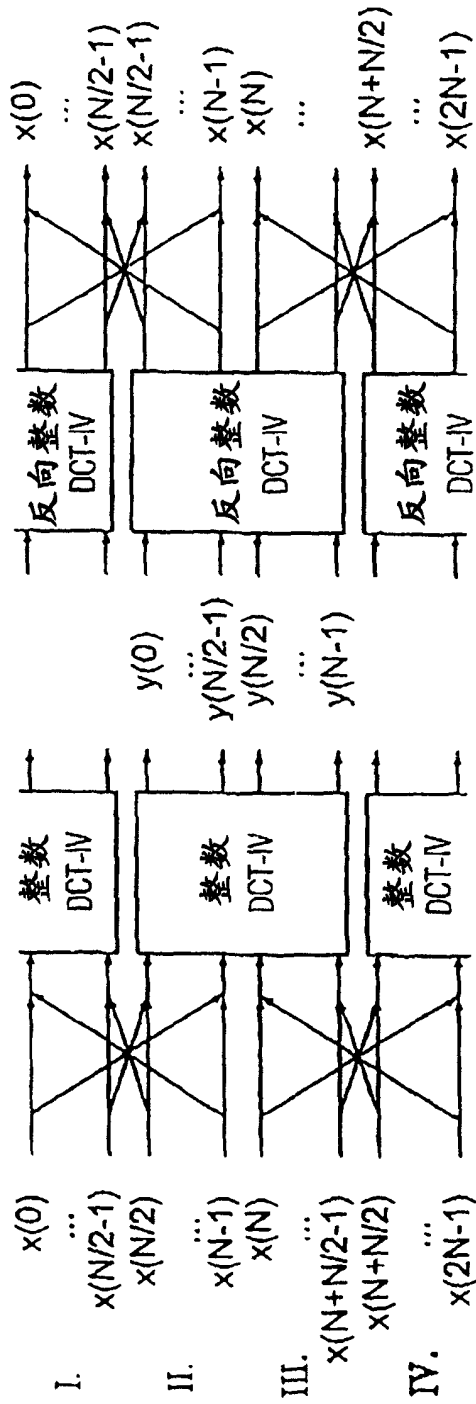


图 2

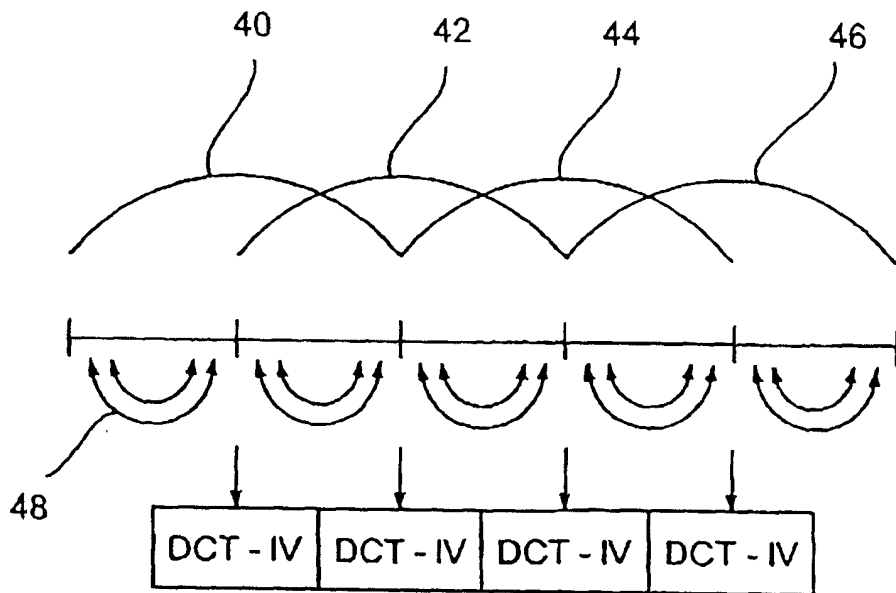


图 3

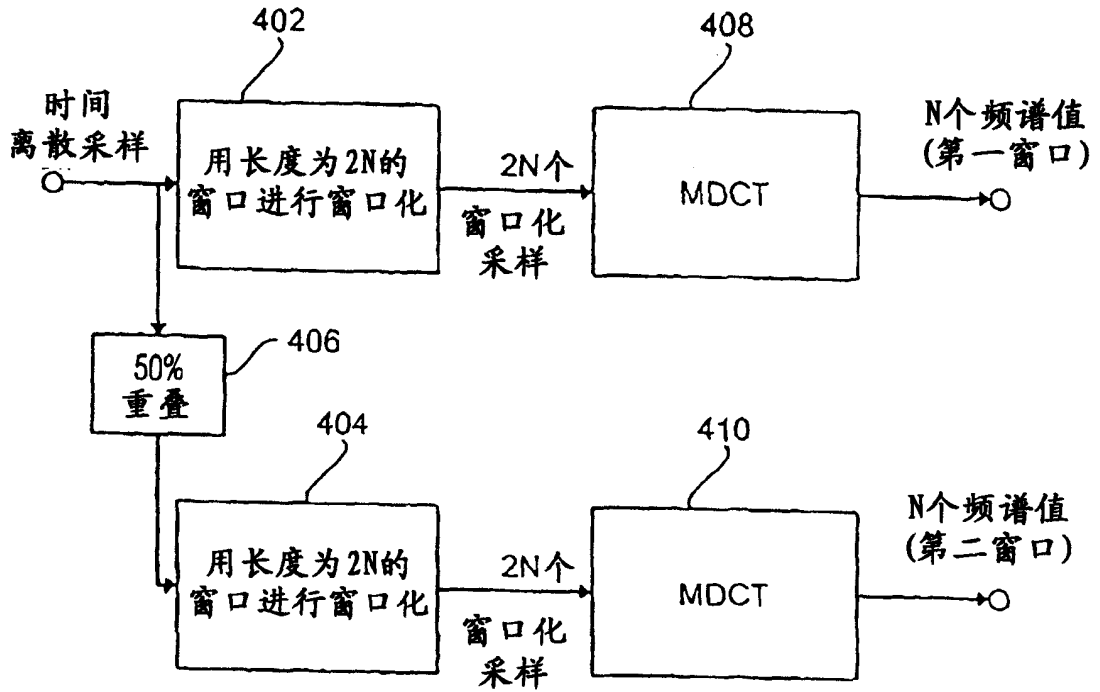


图 4a

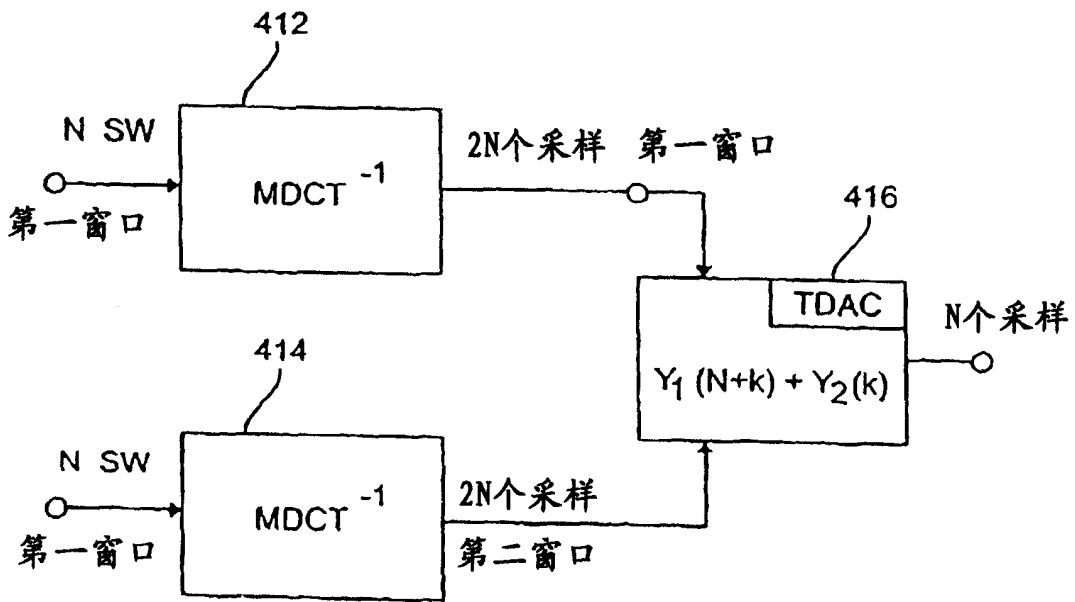


图 4b

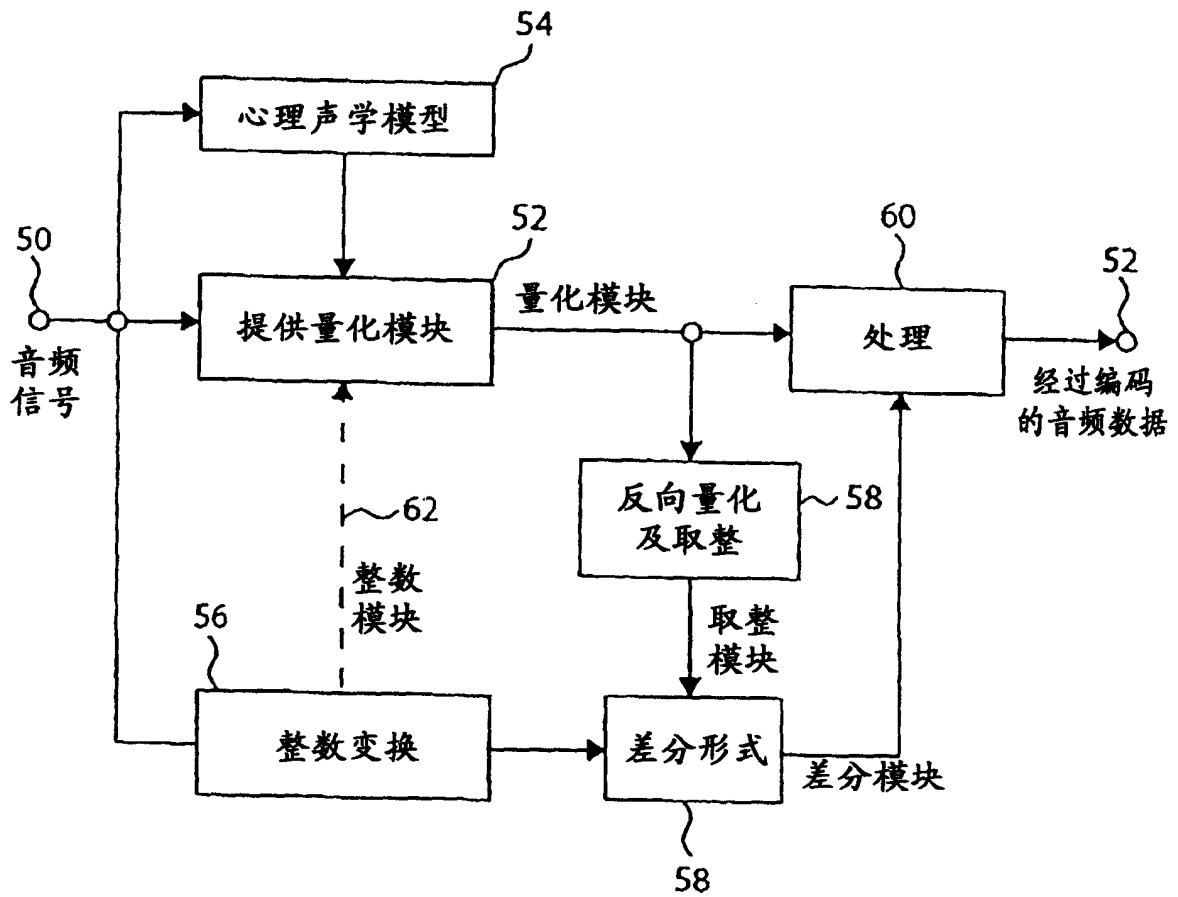


图5

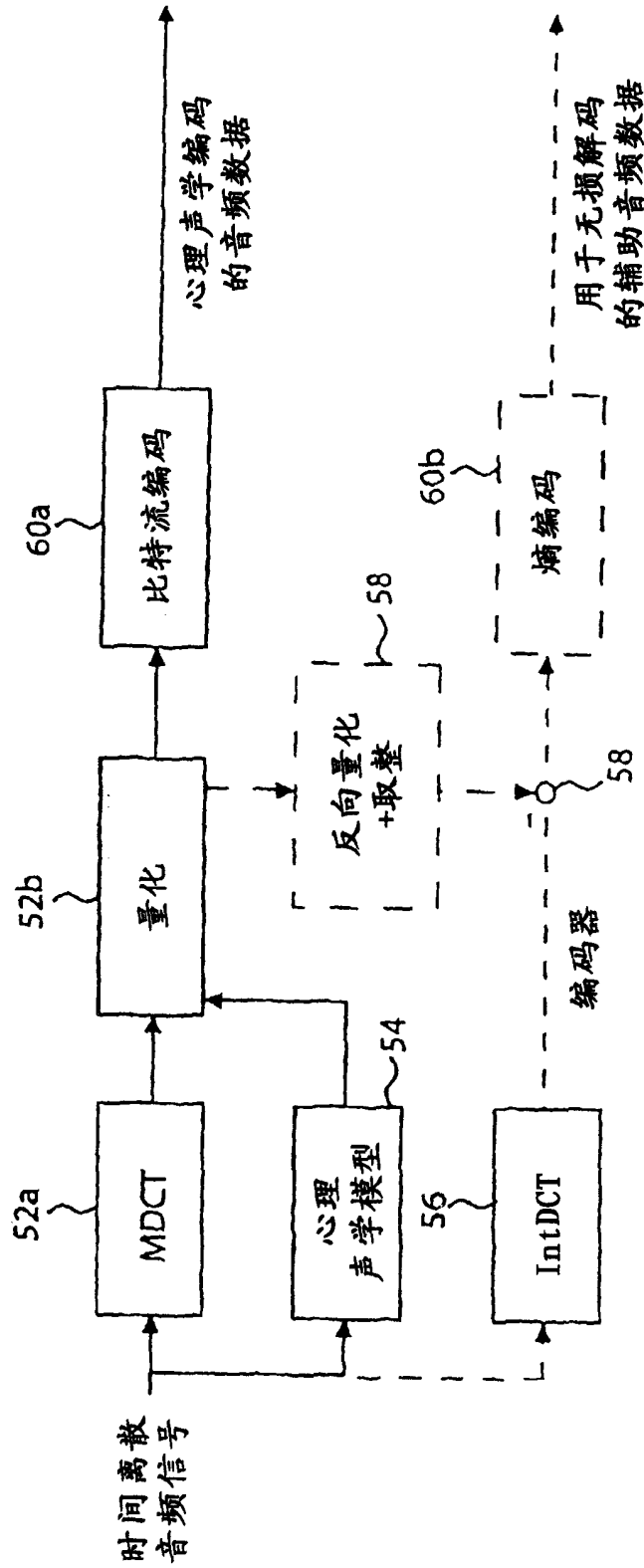


图6

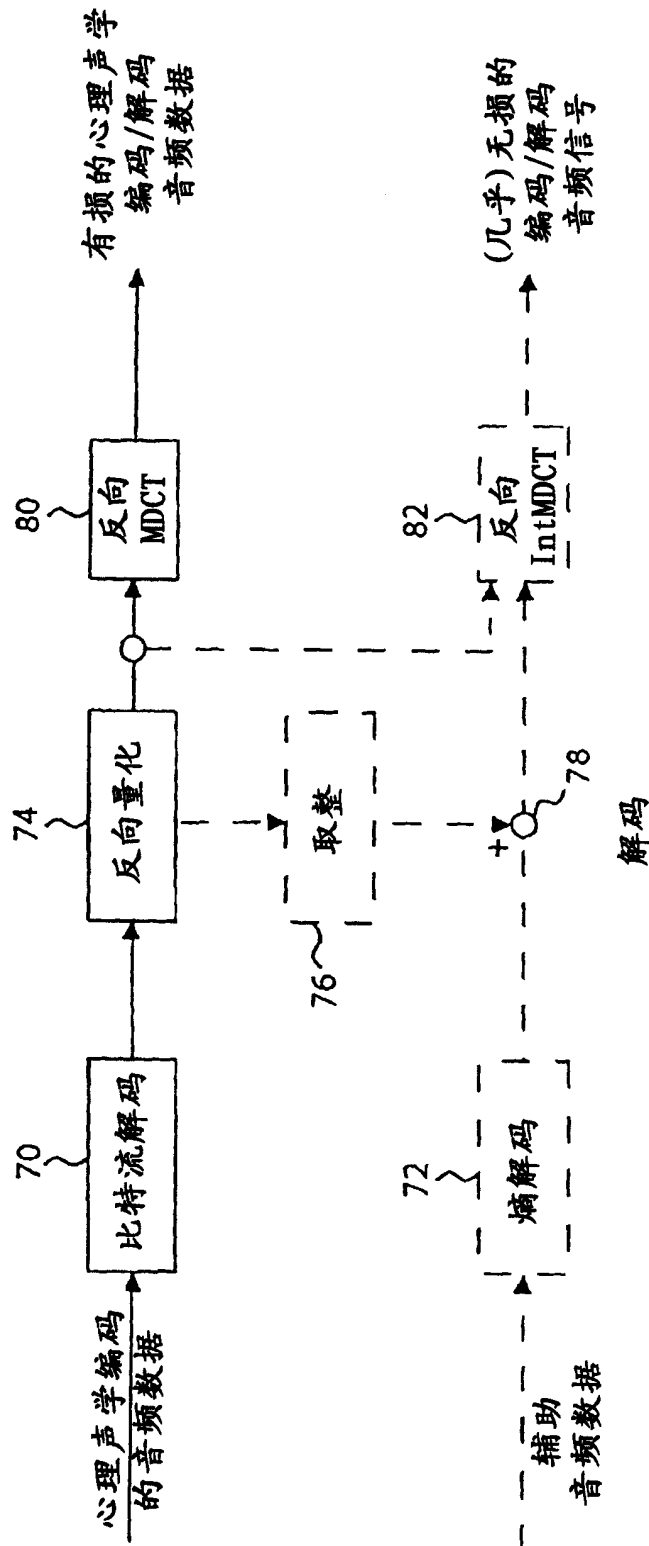


图7

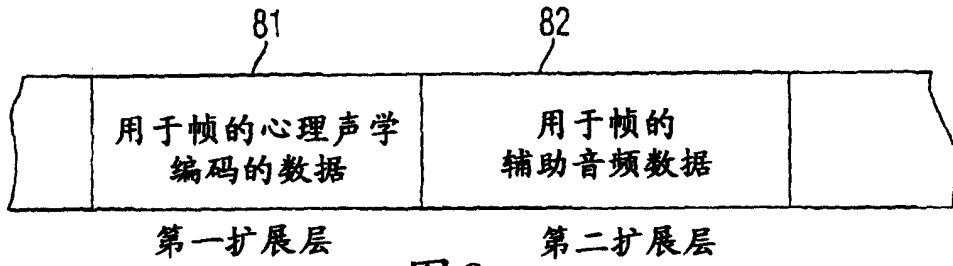


图 8a

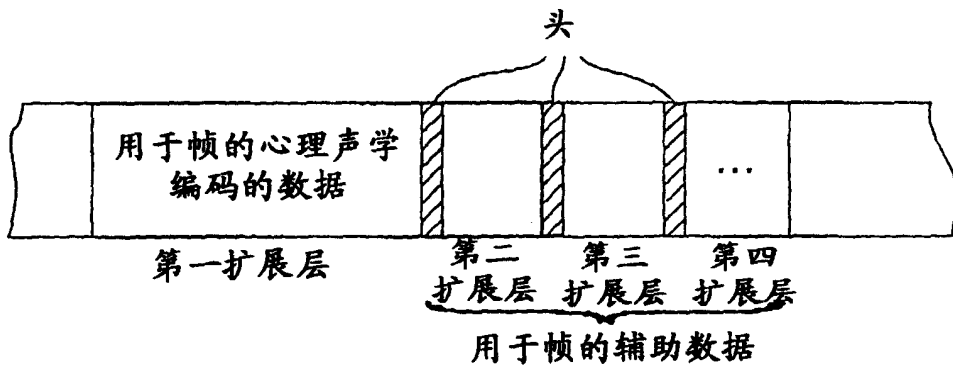


图 8b

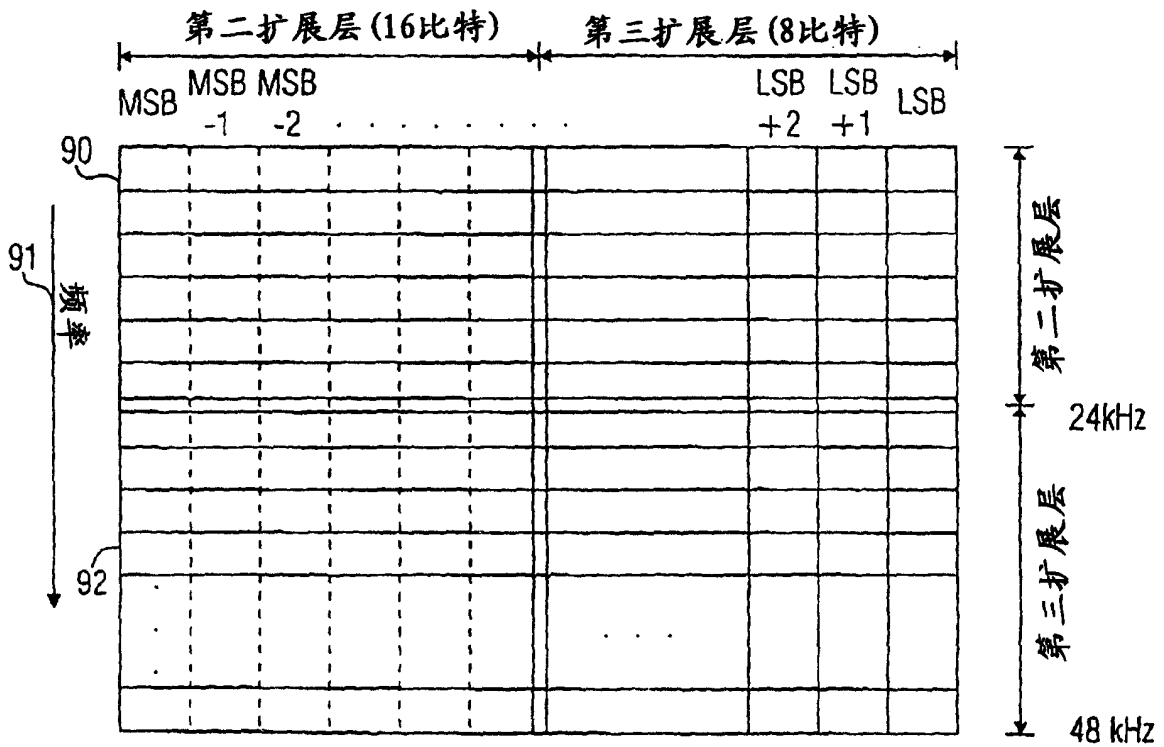


图 9 (差分频谱值)