



[12] 发 明 专 利 说 明 书

[21] ZL 专利号 99809012.3

[45] 授权公告日 2004 年 4 月 14 日

[11] 授权公告号 CN 1146130C

[22] 申请日 1999.5.27 [21] 申请号 99809012.3

[30] 优先权

[32] 1998.5.27 [33] US [31] 09/085,620

[86] 国际申请 PCT/US1999/011898 1999.5.27

[87] 国际公布 WO99/62189 英 1999.12.2

[85] 进入国家阶段日期 2001.1.21

[71] 专利权人 微软公司

地址 美国华盛顿

[72] 发明人 亨里克·S·马瓦

审查员 葛 源

[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

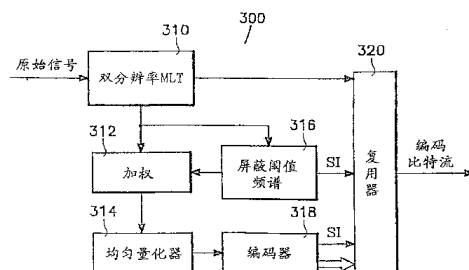
代理人 马 浩

权利要求书 2 页 说明书 26 页 附图 13 页

[54] 发明名称 输入信号处理系统的编码器和屏蔽
频信号量化噪声方法

[57] 摘要

一种处理音频信号和包括可升级音频编码器(300)和解码器的系统和方法。编码器(300)包括一个多分辨率变换处理器(310),例如调制重叠变换(MLT)变换处理器、加权处理器(312)、均匀量化器(314)、屏蔽阈值频谱处理器(316)、熵编码器(318)和通信设备(320),例如用于复用(组合)从上述组件接收的信号以在单一介质上传输的多路复用器(MUX)。编码器(300)通过分辨率转换、频谱加权和数字编码来编码音频信号。而且,执行数字编码信号参数建模以增强编码。解码器包括用于解码已编码音频信号的与编码器(300)相反的组件。



1. 一种处理输入信号的系统中包括一个存储在系统的计算机可读存储器中的编码器，用于对输入信号进行编码，该编码器包括：

一个多分辨率转换处理器（310），接收输入信号并生成具有转换参数的非均匀调制的重叠双正交变换；和

一个加权处理器（312），具有一个屏蔽阈值频谱处理器（316），能够对变换系数进行频谱加权和部分白化以屏蔽量化噪声。

2. 根据权利要求1的编码器，其中通过以下方式来对变换系数进行部分白化（418）：白化预定的加权函数和使每个加权函数变平以生成最终权值从而使噪声频谱峰值衰减，和利用最终权值作为一个函数来屏蔽量化噪声。

3. 根据权利要求1的编码器，其中变换系数通过以近似于听觉模型的临界频带的一个准对数标度计算的一个屏蔽频谱阈值（1312）来频谱加权（412），多个扩展屏蔽阈值在每个临界频带被计算作为来自相邻临界频带影响的最大值（1314），并调整扩展屏蔽阈值生成一个绝对阈值（1316）。

4. 根据权利要求1的编码器，还包括一个解码编码的输入信号的解码器，其中解码器包括一个反加权处理器，具有一个反屏蔽阈值频谱处理器，它接收解码的信号并去屏蔽量化噪声，和一个反多分辨率转换处理器，它接收去屏蔽的编码信号和编码信号的非均匀调制的重叠双正交转换，以生成一个输出信号作为输入信号的可察觉的透明再生。

5. 一种对输入信号进行编码的方法，该方法在包括处理输入信号的编码器的一个系统中实施，该方法包括：

接收输入信号和计算一个具有频率分量的调制的重叠变换（400）；
组合频率分量以生成具有变换系数的非均匀调制的重叠双正交变换（410）；

计算具有听觉屏蔽能力的加权函数；和

将加权函数应用于非均匀调制的重叠双正交变换的变换系数 (412)。

6. 在一个系统中具有一个生成带有给定带宽并有量化噪声的输入信号的变换处理器，一种用于屏蔽量化噪声的方法，该方法包括：

计算以准对准标度的屏蔽频谱阈值，它近似于听觉模型的临界频带 (1312)；

计算在每个临界频带的多个扩展屏蔽阈值作为一个来自相邻临界频带的作用的最大值 (1314)；和

调整扩展屏蔽阈值以生成一个绝对阈值 (1316)。

7. 根据权利要求 6 的方法，其中计算屏蔽频谱阈值包括定义准对数标度的子带频率上限，通过平均所有子带的信号功率来计算准对数标度的多个频谱功率，和计算一个带内屏蔽阈值 (1312)。

8. 根据权利要求 6 的方法，其中调整扩展屏蔽阈值包括比较扩展屏蔽阈值和模型阈值，其中一个最高值被用于所有子带 (1316)。

9. 根据权利要求 6 的方法，其中计算多个扩展屏蔽阈值包括考虑横向屏蔽穿过临界频带，并通过卷积准对数标度的频谱值与三角衰变来计算最大阈值 (1314)。

10. 根据权利要求 9 的方法，其中计算最大阈值包括计算通过卷积所有 Bark 频谱值与三角衰变产生的一个阈值。

11. 根据权利要求 6 的方法，其中计算在每个临界频带的多个扩展屏蔽阈值作为来自相邻临界频带的作用的最大值包括低频扩展极限初始化为一个由卷积频谱值与一个三角衰变表示的一个最大值，并重复地重设低频扩展极限和低频扩展极限，直到由信号带宽覆盖了理想数目的子带 (1314)。

输入信号处理系统的编码器和屏蔽音频信号量化噪声方法

本发明涉及一种用于压缩数字信号的系统和方法，具体地说，涉及一种能够升级数字音频信号的编码和解码的系统和方法。

现在，数字音频表示在许多应用中是很平常的。例如，音乐光盘（CD）、互联网音频剪辑、卫星电视、数字化视频光盘（DVD）和依靠数字音频技术的电话机（有线的或蜂窝的）。通过使用模数（A/D）转换器将模拟音频信号变换成数字信号实现音频信号的数字表示。然后，数字表示可以被编码、压缩、存储、传输和使用等。如果需要，然后可以使用数模（D/A）转换器将数字信号转换回模拟信号。A/D和D/A转换器定期抽样模拟信号，通常以下述标准频率之一：8 kHz，用于电话机、互联网、视频会议；11.025 kHz，用于互联网、只读光盘驱动器；16 kHz，用于视频会议、远距离音频广播、互联网、未来电话机；22.05 kHz，用于只读光盘驱动器、互联网；32 kHz，用于只读光盘驱动器、视频会议、ISDN音频；44.1 kHz，用于音频光盘；和48 kHz，用于播音室音频再生。

典型地，如果在转换之后编码或压缩音频信号，由A/D生成的原始比特通常以每音频抽样16比特被格式化。对于音频光盘，例如，原始比特速率是 $44.1\text{ kHz} \times 16\text{ 比特/抽样} = 705.6\text{ kbps}$ （千比特/秒）。对于电话机，原始比特速率是 $8\text{ kHz} \times 8\text{ 比特/抽样} = 64\text{ kbps}$ 。对于音频光盘，在存储容量大约是700兆字节（5600兆比特）的地方，可以存储原始比特，并且不需要压缩。然而，小型光盘仅能存储大约140兆字节，所以需要大约4:1的压缩以将30分钟到1小时的音频装入2.5英寸小型光盘。

对于互联网电话和大多数其它应用，原始比特速率对于大多数的当前信道容量太高。因而，使用能够有效压缩的高效编码/解码器（通常称作编码器/解码器或编解码器）。例如，对于互联网电话机，

原始比特速率是 64 kHz, 但所需的信道速率在 5 至 10 kbps 之间变化。因此, 编解码器需要以 5 至 15 之间的因子压缩比特速率, 带有最小的可觉察的音频信号质量损失。

随着当前处理芯片的发展, 编解码器可以以专用硬件, 通常用可编程数字信号处理器 (DSP) 芯片实现, 或者以通用计算机软件来实现。因此, 希望使编解码器可以实现, 例如: 1) 低计算复杂性 (编码复杂性对于所存储的音频通常不是问题); 2) 优良的再生保真度 (不同的应用将具有不同的质量要求); 3) 对信号变化的强壮性 (音频信号可以是清楚的话音、嘈杂的话音、多个谈话者、音乐等等, 编解码器可以处理的这种信号的范围越广越好); 4) 低延迟 (在诸如电话机和视频会议的实时应用中); 5) 可升级性 (适应不同信号抽样速率和不同信道容量的容易程度 - 尤其希望编码之后的可升级性, 即转换成不同抽样速率或信道速率而不需要重新编码); 和 6) 压缩域中的信号修改 (如果编解码器允许在压缩域中处理, 或者至少不需要完全解码和重新编码, 诸如多个信道混合、干扰压缩和其它的操作可以更快)。

当前, 商业系统使用多种不同的数字音频技术。一些例子包括: ITU-T 标准: G.711, G.726, G.722, G.728, G.723.1 和 G.729; 其它电话标准: GSM、半速率 GSM、蜂窝 CDMA (IS-733); 高保真音频: 杜比 AC-2 和 AC-3、MPEG LII 和 LIII、索尼小型光盘; 互联网音频: ACELP-Net、DolbyNet、PictureTel Siren、RealAudio; 和军事应用: LPC-10 和 USFS-1016 声码器。

然而, 当前的这些编解码器具有多种限制。即, 当前编解码器的计算复杂性不够低。例如, 当编解码器被集成在一个操作系统中时, 希望使编解码器以低 CPU 使用与其它应用同时运行。另一问题是中等延迟。希望使编解码器允许整个音频获取/重放系统以低于 100 毫秒的延迟操作, 例如, 能够实时通信。

另一问题是对信号变换的强壮性级别。希望使编解码器不仅能够处理清楚话音, 而且能够处理被混响、办公室噪声、电子噪声和背景

音乐等影响的话音，并且还能够在音乐、拨号音和其它声音。而且，大多数现有编解码器的缺点是它们有限的可升级性和支持信号抽样频率和信道数据速率的狭窄范围。例如，多种当前应用通常需要支持多个不同的编解码器。这是因为多种编解码器被设计仅能处理某一范围的抽样速率。一种相关的期望是使编解码器可以允许修改抽样或数据速率而不需要重新编码。

另一问题是多方电话会议，服务器必须混合来自各参与者的音频信号。多个编解码器要求在混合之前解码所有数据流。需要的是编解码器支持在编码或压缩域中混合而不需要在混合之前解码所有数据流。

还有一个问题出现在与信号增强功能的集成中。例如，由当前编解码器使用的音频路径在编解码器处理之前可以包括信号增强模块。举一个例子，在免提电话会议中，来自扬声器的信号被麦克风捕获，被本地人员的话音干扰。因此，通常使用回波消除算法来消除扬声器对麦克风的反馈。其它增强操作部分可以包括自动增益控制、降噪器。那些增强操作部分导致将被增加给编码/解码延迟的处理延迟。因而，所需要的是能够将增强处理和编解码器较简单集成的编解码器，以便可以执行所有这些信号增强而不产生除编解码器延迟之外的任何延迟。

与编解码器相关的另一个问题是缺少对比特和分组损失的强壮性。在大多数实际实时应用中，通信信道不能免于错误。无线信道可能具有明显的误比特率，并且分组交换信道（例如互联网）可能具有明显的分组损失。因而，所需要的是编解码器允许具有小信号降级的压缩比特流损失，例如最高 5%。

不管上述系统和方法的优点如何，它们不能实现本发明的优点。

为了克服上述现有技术中的限制，并克服其它的限制，阅读和理解本发明之后这将变得很明显，在使用新编码器/解码器（编解码器）能够可升级地编码和解码音频信号的一种系统和方法中实施本发明。

本发明的编解码器系统包括一个编码器和一个解码器。编码器包

括多分辨率变换处理器，例如调制重叠变换（MLT）变换处理器、加权处理器、均匀量化器、屏蔽阈值频谱处理器、熵编码器和一个通信设备，例如复用器（MUX），用于复用（组合）从上述部件接收的信号以在单一介质上传输。解码器包括与编码器相反的部件，例如反多分辨率变换处理器、反加权处理器、反均匀量化器、反屏蔽阈值频谱处理器、反熵编码器和解复用器。使用这些部件，本发明能够执行分辨率转换、频谱加权、数字编码和参数建模。

本发明的一些特征和优点包括低计算复杂性。当本发明的编解码器集成在一个操作系统中时，它可以以低 CPU 使用率与其它应用同时运行。这个编解码器允许整个音频捕获/回放系统以低于例如 100 毫秒的延迟操作从而能够进行实时通信。这个编解码器具有对信号变化的高级别强壮性，并且它不仅能够处理清楚话音，能够处理受混响、办公室噪声、电噪声、背景音乐等的影响下降的话音，而且能够处理音乐、拨号音和其它声音。另外，该编解码器能够升级，并支持大范围的信号抽样频率和信道数据速率。一个相关特征是这个编解码器允许修改抽样或数据速率而不需要重新编码。例如，这个编解码器可以将 32 kbps 数据流转换成 16 kbps 数据流而不需要完全解码和重新编码。这使服务器能够仅存储高保真度的音频剪辑，在需要时将它们迅速转换成较低的保真度。

而且，对于多方电话会议，这个编解码器支持在不需要在混合之前解码所有数据流的情况下在编码或压缩域中的混合。这显著地影响服务器可以处理的音频流数量。而且，这个编解码器能够以这种方式进行增强处理的较简单的集成，可以在没有除编解码器延迟之外的任何延迟的情况下执行信号增强。而且，这个编解码器的另一特征是它对比特和分组损失的强壮性。例如，在大多数实际实时应用中，通信信道无法免于错误。因为无线信道可能具有明显的误比特率，分组交换信道（例如互联网）可能具有明显的分组损失，这个编解码器允许压缩比特流损失，例如最高 5%，较低地降低信号质量。

通过研究下述本发明的详细说明和附图及权利要求书，本发明的

上述和其它特征和优点以及对其更完整的理解将更明显。

现在参考附图，其中在所有附图中相同的参考数字表示相应的部件：

图 1 是一个方框图，表示用于执行本发明的一种设备；

图 2 是一个普通方框/流程图，表示根据本发明用于编码/解码音频信号的一种系统和方法；

图 3 是一个总结构方框图，表示根据本发明用于编码音频信号的一个系统；

图 4 是一个总流程图，表示根据本发明用于编码音频信号的方法；

图 5 是一个普通方框/流程图，表示根据本发明用于编码音频信号的一个系统；

图 6 是一个普通方框/流程图，表示根据本发明用于编码音频信号的一个系统；

图 7 是一个流程图，表示根据本发明的一种调制重叠变换；

图 8 是一个流程图，表示根据本发明的一种调制重叠双正交变换；

图 9 是一个简化方框图，表示根据本发明的一种不均匀调制重叠双正交变换；

图 10 表示不均匀调制重叠双正交变换分析基本函数的一个例子；

图 11 表示不均匀调制重叠双正交变换分析基本函数的另一个例子；

图 12 是一个流程图，表示根据本发明用于执行分辨率转换的系统和方法；

图 13 是一个流程图，表示根据本发明用于执行加权函数计算和部分白化的一种系统和方法；

图 14 是一个流程图，表示根据本发明用于执行简化 Bark 阈值计算的一种系统和方法；

图 15 是一个流程图，表示根据本发明用于执行熵编码的一种系统和方法；

图 16 是一个方框图，表示根据本发明用于执行参数建模的一种系统和方法。

在本发明的下述说明中，参考构成本发明一部分的附图，利用在其中可以实施本发明的一个具体例子的例图进行图示。应当理解在不脱离本发明范围的情况下可以使用其它实施例，并可以进行结构上的改变。

引言

在许多现代音频编码标准中使用变换或子带编码器，通常以 32 kbps 和更高的比特率，和以 2 比特/抽样或更高。在低速率上，大约和低于 1 比特/抽样，诸如 G.729 和 G.723.1 的话音编解码器在电话会议应用中使用。这种编解码器以清楚的话音生成模型为基础，因此它们的性能随着其它信号，例如多个说话者，噪声环境，尤其是音乐信号而快速地降低。

随着调制解调器可用速率的提高，许多应用可以为窄带（3.4 kHz 带宽）音频负担高达 8-12 kbps，并且对于更高保真度的材料可以是更高的速率。这提高了在对信号变化更强壮的编码器上的兴趣，该编码器速率例如类似或略高于 G.729。

本发明是带有转换编码器的编码器/解码器系统（编解码器），所述转换编码器可以以合理的质量工作在低到 1 比特/抽样（例如以 8 kHz 抽样的 8 kbps）的速率上。为了在清晰的话音条件下提高性能，使用频谱加权和带有参数建模的行程长度和熵编码器。结果，改善了有声语音的周期频谱结构的编码。

本发明为包括语音的准周期信号产生改善的性能。根据少量参数计算量化参数表，考虑了在不增加量化表存储的情况下高度的适用性。为了提高瞬变信号的性能，本发明使用带有可变分辨率没有输入窗口切换的不均匀调制重叠双正交变换。实验结果表明本发明可以被

用于以接近于 1 比特/抽样的速率的高质量信号再生，以 2 比特/抽样的准透明再生，和以 3 比特/抽样或者更高的速率的可觉察的透明再生。

示范操作环境

图 1 和下述讨论将用于提供在其中可以实施本发明的合适计算环境的简要说明。尽管不需要，但将在计算机可执行指令，例如由个人计算机执行的程序模块的普通环境下描述本发明。通常，程序模块包括执行特殊任务或实现特定摘要数据类型的例程序、程序、对象、组件和数据结构等。而且，本领域的技术人员将认识到可以用其它计算机系统配置实现本发明，包括手持式设备、多处理器系统、基于微处理器或可编程用户电子设备、网络个人计算机、小型计算机、大型计算机等。本发明也可以在由通过通信网络链接的远程处理设备执行任务的分布式计算环境中实施。在分布式计算机环境中，程序模块可以被设置于本地和远程存储设备上。

参见图 1，用于实现本发明的示范性系统包括形式为常规个人计算机 100 的通用计算设备，包括处理单元 102、系统存储器 104 和将包括系统存储器 104 的各种系统组件连接到处理单元 102 的系统总线 106。系统总线 106 可以是多种总线类型中的任何一种总线，包括存储器总线或存储器控制器、外围总线和使用多种总线结构中的任何一种的本地总线。系统存储器包括只读存储器（ROM）110 和随机访问存储器（RAM）112。在 ROM 110 中存储基本输入/输出系统 114（BIOS），包含协助在个人计算机 100 中在组件之间传输信息的基本例程序，例如在启动时。个人计算机 100 还包括用于读取和写入硬盘的硬盘驱动器 116，用于读取或写入活动磁盘 120 的磁盘驱动器 118，和用于读取或写入诸如 CD-ROM 或其它光介质的活动光盘 124 的光盘驱动器 122。硬盘驱动器 116、磁盘驱动器 128 和光盘驱动器 122 分别通过硬盘驱动器接口 126、磁盘驱动器接口 128 和光盘驱动器接口 130 连接到系统总线 106。驱动器和它们的相关计算机可读介

质为个人计算机 100 提供计算机可读指令、数据结构、程序模块和其它数据的非易失性存储。尽管在此所述的示范性环境使用硬盘、活动磁盘 120 和活动光盘 130, 本领域的普通技术人员将理解可以存储可由计算机访问的数据的其它类型的计算机可读介质也可以在示范性操作环境中使用, 例如磁带、快闪存储卡、数字视频光盘、贝努里盒式磁盘、随机访问存储器 (RAM)、只读存储器 (ROM) 等。

多个程序模块可以被存储在硬盘、磁盘 120、光盘 124、ROM 110 或 RAM 112 上, 包括操作系统 132、一个或多个应用程序 134、其它程序模块 136 和程序数据 138。用户可以通过诸如键盘 140 和指示设备 142 的输入设备将命令和信息输入个人计算机 100。其它输入设备 (未图示) 可以包括麦克风、控制杆、游戏小键盘、卫星反射器、扫描仪等。这些或其它输入设备通常通过连接到系统总线 106 的串行接口 144 连接到处理单元 102, 但是也可以通过其它接口连接, 例如并口、游戏口或通用串行总线 (USB)。监视器 146 或其它类型的显示设备也通过诸如视频适配器 148 的接口连接到系统总线 106。除了监视器 146 之外, 个人计算机通常包括其它外围输出设备 (未图示), 例如扬声器和打印机。

个人计算机 100 可以使用到诸如远程计算机 150 的一个或多个远程计算机的逻辑连接工作在网络环境中。远程计算机 150 可以是另外一个个人计算机、服务器、路由器、网络 PC、对等设备或其它普通网络节点, 并且一般包括上述与个人计算机 100 相关的多个或所有组件, 尽管仅在图 1 中图示存储器存储设备 152。图 1 所示的逻辑连接包括局域网 (LAN) 154 和广域网 (WAN) 156。这些网络环境在办公室、大企业计算机网络、内部网和因特网中是常见的。

当在 LAN 网络环境中使用时, 个人计算机 100 通过网络接口或适配器 158 连接到本地网 154。当在 WAN 网络环境中使用时, 个人计算机 100 一般包括调制解调器 160 或其它装置, 用于在诸如因特网的广域网 156 上建立通信。可以是内置或外置的调制解调器 160 通过串口接口 144 连接到系统总线 106。在网络环境中, 与个人计算机 100

或它的一部分相关描述的程序模块可以被存储在远程存储器存储设备中。显然所示的网络连接是示范性的，可以使用在计算机之间建立通信链路的其它装置。

概述

图 2 是一个普通方框/流程图，说明根据本发明用于编码/解码一个音频信号的系统和方法。首先，由模数（A/D）转换器 210 接收和处理一个信号源的模拟音频输入信号。A/D 转换器 210 生成原始数据比特。原始数据比特被发送给数字编码器 212，并根据本发明处理以生成一个编码比特流（下文提供对编码器的详细说明）。编码比特流被使用、存储、传输等（方框 204），然后发送给数字编码器 216 和处理以再生原先的原始数据比特。数模（D/A）转换器 218 接收原始数据比特以转换成输出音频信号。所生成的输出音频信号与输入音频信号基本匹配。

图 3 是一个总结构方框图，图示根据本发明用于编码音频信号的一个系统。本发明的编码器 300（图 2 的编码器 212）包括多分辨率变换处理器 310、加权处理器 312、均匀量化器 314、屏蔽阈值频谱处理器 316、编码器 318 和通信设备 320。

多分辨率变换处理器 310 最好是双分辨率调制重叠变换（MLT）变换处理器。变换处理器接收初始信号并根据初始信号生成变换系数。加权处理器 312 和屏蔽阈值频谱处理器 316 为尽可能地屏蔽量化噪声执行频谱加权和部分白化。均匀量化器 314 用于将连续值变换成离散值。编码器 318 最好是用用于编码变换系数的熵编码器。通信设备 320 最好是复用器（MUX），用于复用（组合）从上述组件接收的信号以在单一介质上传输。

解码器（未图示）包括编码器 300 的相反组件，例如反多分辨率变换处理器（未图示）、反加权处理器（未图示）、反均匀量化器（未图示）、反屏蔽阈值频谱处理器（未图示）、反编码器（未图示）和反 MUX（未图示）。

组件概述

图 4 是一个总流程图，表示根据本发明用于编码音频信号的方法。将讨论图 7-16 中操作的具体细节。一般来说，首先，执行一个 MLT 计算（方框 400）以生成变换系数，继之以修改 MLT 系数（方框 410）的分辨率切换（方框 405）。分辨率转换被用于提高瞬变信号的性能。

然后，通过下述步骤执行频谱加权（方框 412）：a）根据下面将要描述的本发明的听觉屏蔽技术加权变换系数（方框 414）；b）计算简化 Bark 阈值频谱（方框 416）；c）执行加权函数的部分白化（方框 418）；和 d）执行标量量化（方框 420）。根据本发明执行频谱加权以尽可能多的屏蔽量化噪声。目的是生成尽可能接近清晰的的重建信号。

第三，执行编码和参数建模（方框 422），通过建立由编码器使用的概率分布模型（步骤 424），所述编码器例如用于熵编码量化系数的熵编码器（方框 426），然后为量化步长优化执行二进制搜索（方框 428）。标量量化（方框 420）将浮点系数转换成量化系数，它由一组离散数中最接近的值给出。离散值之间的距离等于步长。其中熵编码和参数建模提高清晰话音环境下的性能。熵编码生成在一个消息中用一个码元表示的信息的平均数量，并且是用于生成该消息的概率模型（参数建模）的函数。增加模型的复杂性以便模型可以更好地反映信源码元在始发消息中的实际分布从而减小消息。这个技术能够提高有声话音周期频谱结构的编码。

图 5 是一个总方框/流程图，表示根据本发明用于编码音频信号的系统。图 6 是一个总方框/流程图，表示根据本发明用于解码音频信号的系统。一般来说，输入信号 $x(n)$ 的重叠块由编码器 500 通过非均匀调制重叠双正交转换（NMLBT）510 转换到频域。NMLBT 510 本质上是带有不同分析和合成窗口的调制重叠转换（MLT），其中为了更好的时间分辨率组合高频子带。根据信号频谱，高频子带的组合

可以被打开或关闭，并且一个比特的标记被作为图 6 解码器的辅助信息发送。如下面更加详细的讨论，不修改 NMLBT 分析和合成窗口。

如图 5 所示，利用均匀量化器 512 量化变换系数 $X(k)$ 。在速率失真的意义上，均匀量化器 512 非常接近于最佳，如果它们的输出被例如用行程长度和 Tunstall 编码器 514 进行熵编码（下文进行详细描述）。可以使用矢量量化（VQ），但与熵编码器 514 相比性能增加较小。尽管双 VQ 或其它结构的 VQ 可以用于降低复杂性，它们比标量量化依然复杂得多。

在任意给定比特率上最小化失真的最佳速率分配规则将为子带/转化系数指定相同的步长，生成量化白噪声。这生成最大的信噪比（SNR），但不是最佳听觉质量。加权函数计算 516 在量化之前用 $X(k)/w(k)$ 替换 $X(k)$ ， $k=0,1,\dots,M-1$ ，其中 M 是子带数，通常是 256 和 1024 之间 2 的幂。在图 6 的解码器上，加权通过 $\hat{X}(k) \leftarrow \hat{X}(k)w(k)$ 重新构建的变换系数。因而，量化噪声将跟随通过加权函数定义的频谱 $w(k)$ 。下面的段落描述 $w(k)$ 的详细计算。利用熵编码器 514 熵编码变换系数。执行参数建模并由熵编码器 514 使用结果以增加熵编码器 514 的效率。而且，步长调整 518 被用于调整步长。

可以从图 5 推出图 6 解码器的操作。除了对应于量化变换系数的编码比特之外，图 6 的解码器还需要图 5 所示的辅助信息，以便它可以确定用于反 NMLBT 的单/多分辨率标志、熵解码表、量化步长、和加权函数 $w(k)$ 。

组件细节和操作

再次参考图 3 和图 5，通过诸如重叠转换处理器的转换处理器将输入音频信号分解成频率分量。这时因为尽管诸如离散余弦转换（DCT 和 DCT-IV）的其它转换处理器是用于频域信号分解的有用工具，它们受块人工噪声的影响。例如，由 DCT 和 DCT-IV 转换处理器以一些指定方式：量化、滤波、降噪等处理变换系数 $X(k)$ 。

通过对这些修改后系数施加逆转换获得重建的信号块。当这些重

建的信号块被一起传递以构成重建信号（例如解码音频或视频信号）时，在块边界上将存在不连续性。与此相反，调制重叠转换（MLT）通过将基本函数的长度扩展为块大小的两倍即 $2M$ 来消除这种不连续性。图 7 是表示根据本发明的调制重叠转换的流程图。

通过以下述方式扩展 DCT-IV 函数并将它们乘以一个合适的窗口获得 MLT 的基本函数，

$$a_{nk} = h(n) \cos \left[\left(n + \frac{M+1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right]$$

其中 k 从 0 变化到 $M-1$ ，而 n 从 0 变化到 $2M-1$ 。

因此，最好使用 MLT，因为它们可能产生正交或双正交基础，并可以将信号的短期分解实现为重叠窗口余弦函数的叠加。这种函数为信号的定位频率分解提供比 DCT 或 DCT-IV 更加有效的工具。MLT 是允许理想重建的余弦调制滤波器组的特殊形式。例如，根据信号的 MLT 系数可以精确地恢复信号。而且，MLT 并不具有块人工噪声，即 MLT 提供在其边界上平滑衰变到零的重建信号，避免了沿块边界的不连续性。另外，从速率/失真意义上说，对于各种信号的转换编码，MLT 具有几乎最佳的性能。

具体地说，MLT 基于奇数栈式时域混叠消除（TDAC）滤波器组。一般来说，用于包含输入信号 $x(n)$ ， $n=0,1,2,\dots,2M-1$ 的 $2M$ 个抽样的一个矢量（通过移位输入信号最后 M 个抽样，并将它们与先前获取的 M 个抽样组合来确定）的标准 MLT 转换被转换成包含 M 个系数 $X(k)$ ， $k=0,1,2,\dots,M-1$ 的另一个矢量。该转换可以用标准 MLT 计算来重新定义：

$$X(k) = \sqrt{\frac{2}{M}} \sum_{n=0}^{2M-1} x(n) h(n) \cos \left[\left(n + \frac{M+1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right]$$

其中 $h(n)$ 是 MLT 窗口。

窗口函数主要为降低块效应而使用。例如，使用重叠转换的信号处理（Signal Processing with Lapped Transforms），H.S.Malvar, Boston: Artech 出版社 1992 年，在此引用作为参考，论证了以下述形式通过余弦调制平滑窗口算子获取其基本函数：

$$\begin{aligned} p_a(n,k) &= h_a(n) \sqrt{\frac{2}{M}} \cos \left[\left(n + \frac{M+1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right] \\ p_s(n,k) &= h_s(n) \sqrt{\frac{2}{M}} \cos \left[\left(n + \frac{M+1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right] \end{aligned} \quad (1)$$

其中 $p_a(n,k)$ 和 $p_s(n,k)$ 是用于直接（分析）和反向（合成）转换的基本函数， $h_a(n)$ 和 $h_s(n)$ 分别是分析和合成窗口。时间索引 n 从 0 变化到 $2M-1$ 和频率索引 k 从 0 变化到 $M-1$ ，其中 M 是块大小。MLT 是 TDAC，为其窗口生成具有最大 DC 密度的重叠转换，即：

$$h_a(n) = h_s(n) = \sin \left[\left(n + \frac{1}{2} \right) \frac{\pi}{2M} \right] \quad (2)$$

直接转换矩阵 P_a 在第 n 行 k 列具有项目 $p_a(n,k)$ 。类似地，逆变换矩阵具有项目 $p_s(n,k)$ 。对于信号 $x(n)$ 的 $2M$ 输入抽样的块 x ，通过 $X = P_a^T x$ 计算其变换系数的相应矢量 X 。对于处理变换系数的矢量 Y ，通过 $y = P_s Y$ 给出重建的 $2M$ 抽样矢量 y 。重建矢量 y 与 M 抽样重叠叠加，生成重建信号 $y(n)$ 。

可以比较 MLT 与 DCT-IV。对于信号 $u(n)$ ，其长度 M 正交 DCT-IV 通过下式定义：

$$U(k) \equiv \sqrt{\frac{2}{M}} \sum_{n=0}^{M-1} u(n) \cos \left[\left(n + \frac{1}{2} \right) \left(k + \frac{1}{2} \right) \frac{\pi}{M} \right] \quad (3)$$

构成 DCT-IV 基础的余弦函数的频率是 $(k+1/2)\pi/M$ ，与 MLT 的相同。因此，存在两个转换之间的简单关系。例如对于具有 MLT 系数 $X(k)$ 的信号 $X(n)$ ，它可以被表示为 $X(k)=U(k)$ ，如果 $u(n)$ 涉及 $x(n)$ ， $n=0,1,\dots,M/2-1$ ，通过

$$\begin{aligned} u(n+M/2) &= \Delta_M \{ x(M-1-n)h_a(M-1-n) - x(n)h_a(n) \} \\ u(M/2-1-n) &= x(M-1-n)h_a(n) + x(n)h_a(M-1-n) \end{aligned}$$

其中 $\Delta_M\{\}$ 是 M -抽样（一块）延迟算子。为了说明性的目的，通过组合 DCT-IV 与上述，可以根据标准 DCT-IV 计算 MLT。可以以类似的方式获取一个反向 MLT。例如，如果 $Y(k)=X(k)$ ，即没有任何变换系数（或子带信号）的修改，则级联直接和反向 MLT 处理信号产生 $y(n)=x(n-2M)$ ，其中 M 个延迟抽样来自块算子，另外 M 个抽样来自 MLT 的内部重叠算子（ z^{-M} 算子）。

调制重叠双正交转换

在本发明中，实际的最佳转换是调制重叠双正交转换（MLBT）。图 7 是表示根据本发明的调制重叠双正交转换的流程图。MLBT 是调制重叠转换（MLT）的变型。与 MLT 类似，MLBT 窗口长度是块大小的两倍。它产生最大编码增益，但它的形状针对初始 MLT 正弦窗口进行了稍微修改。为了生成在等式（1）中用公式表示的双正交 MLT，需要放松相同分析和合成窗口的限制。假设一个对称合成窗口，并将双正交条件应用于等式（1），等式（1）生成调制重叠双正交转换（MLBT），如果分析窗口满足一般化条件：

$$h_a(n) = \frac{h_s(n)}{h_s^2(n) + h_s^2(n+M)}, n = 0, 1, \dots, M-1 \quad (4)$$

$$\text{和 } h_a(n) = h_a(2M-1-n)。$$

可以为最大转换编码增益最佳化该窗口，结果是最佳窗口收敛到等式（2）的 MLT 窗口。这允许 MBLT 改善合成基本函数响应的频率选择性，并被用作非均匀 MLT 的标准块（在下文进行详细说明）。MBLT 可以被定义为带有合成窗口的等式（1）的调制重叠转换：

$$h_s(n) = \frac{1 - \cos\left[\left(\frac{n+1}{2M}\right)^a \pi\right] + \beta}{2 + \beta}, n = 0, 1, \dots, M-1 \quad (5)$$

和分析窗口用等式（4）定义。

参数 α 主要控制窗口宽度，而 β 控制它的结果值。MLBT 与 MLT 相比的主要优点是合成函数阻带衰减的增加，以分析函数阻带衰减的降低为代价。

NMLBT 和分辨率切换

典型转换编码器的子带数 M 必须足够大以提供足够的频率分辨率，这经常导致块大小在 20 - 80 毫秒的范围内变化。这导致对瞬变信号很差的响应，带有延续整个块的噪声模式，包括前回声。在这种瞬变信号过程中不需要很好的频率分辨率，因此，减轻该问题的一种方式是为这种声音使用较小的 M 。为调制重叠转换切换块大小并不

困难,但是可能引入附加的编码延迟。一种可选方法是使用分级转换或树形结构滤波器组,类似于离散小波变换。这种分解实现一个新的非均匀子带结构,带有助于高频子带的小块和用于低频子带的大块。分级(或级联)变换具有块之间理想的时域分离,但是很差的频域分离。例如,如果在子带上 QMF 滤波器组继之以 MLT,位于 QMF 过渡带附近的子带可能具有低达 10dB 的阻带抑制,一个在树形结构变换上也发生的问题。

创建一个新的非均匀变换结构以降低 MLT/MLTBT 的振铃人工噪声的可选和最佳方法可以通过修改时间-频率分辨率实现。变换的时间-频率分辨率的修改可以通过将附加变换算子应用于变换系数组以生成一个变换系数的新组合,它生成一个特定的非均匀变换 MLBT (NMLBT)。图 7 是表示根据本发明的非均匀调制重叠双正交变换的简化方框图。

图 8 是表示根据本发明的非均匀调制重叠双正交变换的操作的简化方框图。具体地说,通过线性地组合一些子带系数 $X(k)$ 可以生成一种非均匀 MBLT,并且新子带的滤波器具有降低时间宽度的脉冲响应。一个例子是:

$$X'(2r) = X(2r) + X(2r+1)$$

$$X'(2r+1) = X(2r) - X(2r+1)$$

其中以频率 $(2r+1/2)\pi/M$ 和 $(2r+3/2)\pi/M$ 为中心的子带信号 $X(2r)$ 和 $X(2r+1)$ 被组合以生成两个新的子带信号 $X'(2r)$ 和 $X'(2r+1)$ 。这两个新的子带信号都以 $(r+1)\pi/M$ 为中心,但是一个具有以块左侧为中心的脉冲响应,而另一个具有以块右侧为中心的脉冲响应。因此,我们舍弃频率分辨率来增加时间分辨率。图 9 图示非均匀调制重叠双正交变换合成基本函数的一个例子。

这种通过组合变换系数的分辨率转换的方法的主要优点在于时间分辨率更窄的新的子带信号可以在计算输入信号的 MLT 之后计算。因此,不需要转换 MLT 窗口函数或块大小 M 。它还允许信号增强算子,例如降噪器或回波消除器,以在子带合并算子之前对初始变换/

子带系数进行操作。这允许将这些信号增强器集成到编解码器中。

可以选择地，并且最好是，如果通过因子 4 提高时间分辨率可以实现更好的结果。这导致有效宽度为四分之一块大小的子带滤波器脉冲响应，具有结构：

$$\begin{bmatrix} X'(4r) \\ X'(4r+1) \\ X'(4r+2) \\ X'(4r+3) \end{bmatrix} = \begin{bmatrix} a & a & a & a \\ b & c & -c & -b \\ -b & c & c & -b \\ -a & a & -a & a \end{bmatrix} \begin{bmatrix} X(4r) \\ X(4r+1) \\ X(4r+2) \\ X(4r+3) \end{bmatrix}$$

其中参数的最佳选择是 $a=0.5412$, $b=\sqrt{1/2}$, $c=a^2$, $r=M_0, M_0+1, \dots$, 并且 M_0 一般被设置为 $M/16$ (这意味着分辨率转换被应用于子带的 75% - 从频率 0.25π 到 π)。图 10 和图 11 表示与这种结构对应的合成基本函数的曲线。可以看出时间分离不理想，但是它降低了瞬变信号的错误扩散。

上述子带组合矩阵的自动转换可以通过分析输入块波形在编码器上进行。如果块中的概率级别显著地变化，组合矩阵被打开。转换标志作为辅助信息被发送给接收机，所以它可以使用逆 4×4 算子来恢复 MLT 系数。一种可选的转换方法是分析 MLT 系数 $X(k)$ 之间的功率分配，并在检测到高频类噪声模型时转换组合矩阵。

图 12 是表示根据本发明执行分辨率转换的最佳系统和方法的流程图。如图 12 所示，在每个块上确定分辨率转换，并将辅助信息的一个比特发送给解码器以通知开关是接通还是断开。在最佳实施中，当用于给定块的高频能量超过低频能量预定阈值时，编码器将开关接通。基本上，编码器通过测量在高频和低频上的信号功率来控制分辨率转换。如果高频功率 (PH) 与低频功率 (PL) 之比超过预定阈值，使用步骤 2 中所述的子带组合矩阵，如图 12 所示。

频谱加权

图 13 是表示根据本发明用于执行加权函数计算和部分白化的系统和方法的流程图。再次参考图 3、图 5 和图 13，图示一种用于执行频谱加权的简单技术。根据本发明，可以执行频谱加权以尽可能多地

屏蔽量化噪声。目的是生成尽可能清晰的重建信号，即无法区分解码信号和原始信号。这可以通过用函数 $w(k)$ 加权变换系数来实现，该函数依靠人耳的屏蔽特性。这种加权的意义是形成人耳觉察程度最低的量化噪声，因而屏蔽了量化噪声。而且，简化听觉加权函数的计算以避免通常使用的耗费时间的卷积。

对于给定的输入频谱 $\{X(k)\}$ ，加权函数 $w(k)$ 理论上沿着一个听觉屏蔽阈值曲线。最好以 Bark 标度计算屏蔽阈值。Bark 标度(scale)是模拟人耳临界频带的准对数标度。在高编码率上，例如 3 比特/抽样，对于所有的 Bark 子带，结果量化噪声可能低于量化阈值从而生成清晰的重建。然而，在低编码速率上，例如 1 比特/抽样，在屏蔽阈值下隐藏所有的量化噪声是困难的。在这种情况下，最好防止量化噪声在所有子带中被抬高得高于屏蔽阈值相同的分贝数 (dB)，因为低频未屏蔽噪声通常更加有害。这可以通过用新的函数 $W(k)^\alpha$ 替换初始加权函数 $W(k)$ 来实现，其中 α 是通常被设置得低于 1 的参数，以建立加权函数的部分白化。

一般来说，参考图 13、图 3、图 4 和图 5，图 13 表示听觉阈值曲线的简单计算，以及用于计算步长的部分白化效应。图 13 是图 3 的方框 312 和 316、图 4 的方框 414、416、418 和图 5 的方框 516 的详细方框图。参考图 13，在 MLT 计算和 NMLBT 修改之后，变换系数 $X(k)$ 首先被平方模块接收以将变换系数平方（方框 1310）。接着，阈值模块计算 Bark 频谱阈值（方框 1312），该阈值由扩展模块用于执行 Bark 阈值扩展（方框 1314）和生成听觉阈值。然后，调整模块将听觉阈值调整为绝对阈值以生成一个理想加权函数（方框 1316）。最后，执行部分白化效应以便将理想加权函数提高到第 α 功率从而生成最终的加权函数（方框 1318）。

具体地说，平方模块生成 $P(i)$ ，在第 i 频带上的瞬时功率，它由阈值模块接收用于计算屏蔽阈值 $w_{MT}(k)$ ，（如图 13 的方框 1310 所示）。这可以下述方式实现：一开始定义 Bark 频谱频率上限 $Bh(i)$ ， $i=1, 2, \dots, 25$ （可以使用的常规数学策略），以便 Bark 子带上限是（单

位: Hz):

$Bh=[100\ 200\ 300\ 400\ 510\ 630\ 770\ 920\ 1080\ 1270\ 1480\ 1720\ 2000];$
 $Bh=[Bh\ 2320\ 2700\ 3150\ 3700\ 4400\ 5300\ 6400\ 7700\ 9500\ 12000$
 $15500\ 22200];$

接着, 通过平均落入第 i Bark 频带的所有子带的信号功率来计算第 i Bark 频谱功率 $Pas(i)$ 。然后通过 $Tr(i)=Pas(i)-Rfac$ 计算带内屏蔽阈值 $Tr(i)$ (所有数值的单位都是 dB)。参数 $Rfac$, 它最好被设置为 7dB, 确定带内屏蔽阈值级。这可以通过一个数学循环程序来实现以生成 Bark 功率频谱和 Bark 中央阈值。

如图 13 的方框 1314 所示, 随后计算一个简化 Bark 阈值频谱。图 14 图示根据本发明的简化 Bark 阈值计算。最好通过考虑穿越临界频带的横向屏蔽来计算“频谱 Bark 阈值”。即, 并不如前面的方法所建议的利用矩阵算子执行一个完整卷积, 本发明使用通过卷积所有 Bark 频谱值与三角衰变生成的曲线中的最大阈值曲线。三角衰变对于左侧是 -25dB/Bark (扩展到低频) 和对于右侧是 +10dB/Bark (扩展到高频)。用于 Bark 频谱阈值扩展的本发明的方法具有复杂度 $O(Lsb)$, 其中 Lsb 是由信号带宽覆盖的 Bark 子带数目, 而前面的方法一般具有复杂度 $O(Lsb^2)$ 。

如图 13 的方框 1316 所示, 然后通过比较扩展 Bark 阈值与绝对 Fletcher-Munson 阈值, 并将两者中的较高者用于所有的 Bark 子带, 调整听觉阈值。这可以通过一个简单程序来实现, 例如通过调整考虑绝对屏蔽的阈值。在一个程序中, 阈值向量 (多达 25 个每块) 被量化成预定精度级别, 一般设置为 2.5dB, 并且 2 至 4 比特/阈值被差分编码。

对于加权函数的部分白化, 如图 13 的方框 1318 所示, 以较低的速率, 例如 1 比特/抽样, 不可能将所有的量化噪声隐藏在屏蔽阈值之下。在这种特定情况下, 最好不要将量化噪声在所有子带提高得高于屏蔽阈值相同的 db 值, 因为低频未屏蔽噪声通常更加有害。因此, 假设 $w_{MT}(k)$ 是如上计算的加权, 本发明的编码器使用最终的加权:

$$w(k) = [w_{MT}(k)]^\alpha$$

其中 α 是可以从低速率上的 0.5 变化到高速率上的 1 的参数, 并最好使用屏蔽阈值的分数幂。在先前的听觉编码器中, 在所有频率上相等地将量化噪声提高得高于屏蔽阈值, 如同降低比特率。相反, 使用本发明, 例如, 可以将部分白化参数 α 设置为 0 至 1 之间的一个数 (优选地, $\alpha = 0.5$)。这导致噪声频谱在其原先较低的频率上被提得很高。换句话说, 当 $\alpha < 1$ 时噪声频谱峰值被衰减。

接着, 用于表示 $w'(k)$ 的辅助信息 (side frequency) 的数值取决于抽样频率 f_s 。例如, 对于 $f_s = 8$ kHz, 需要大约 17 Bark 谱值, 对于 $f_s = 44.1$ kHz, 需要大约 25 Bark 谱值。假设 -10dB/Bark 频带的到更高子带的段间扩展和用 2.5 dB 精度的差分编码, 需要大约 3 比特/Bark 系数。可以利用标量量化器量化加权变换系数 (从连续值变换成离散值)。

具体地说, 对于标量量化, 最终加权函数 $w(k)$ 确定将最不容易觉察的量化噪声的频谱形状, 根据上面讨论的模型。因此, 每个子带频率系数 $X(k)$ 应当用与 $w(k)$ 成比例的步长量化。一个等效程序将所有的 $X(k)$ 除以加权函数, 然后对所有系数 $X(k)$ 以相同步长进行均匀量化。一个典型的实现是执行下述:

```
Xr=round(X/dt); %quantize (量化)
```

```
Xqr=(Xr+Rqnoise)*dt; %scale back, adding pseudo-random noise (量化回归, 加入伪随机噪声)
```

其中 dt 是量化步长。向量 $Rqnoise$ 由均匀分布在 $[-\gamma, \gamma]$ 之间的伪随机变量组成, 其中 γ 是最好在 0.1 和 0.5 倍量化步长 dt 之间选择。通过相加少量噪声和重建系数 (一个解码器操作), 可以减少因频谱分量丢失导致的人工噪声。这可以被称为拦动、伪随机量化或噪声填充。

编码

在信息理论中典型的离散信源编码问题是以最经济的编码表示信

源码元的问题。例如，假设信源在每个瞬间 i 上发送码元 s_i ，并且码元 s_i 属于字母表 Z 。并且，假设码元 s_i 和 s_i 是统计独立的，具有概率分布 $\text{Prob}\{s_i=z_n\}=P_n$ ，其中 $n=0, 1, \dots, N-1$ ，并且 N 是字母表大小，即可能的码元数。因此，编码设计问题是发现利用信道码元通常是比特来表示码元 s_i 的方法。

可以使用普通编码为每个可能码元值 z_n 分配一个 M 比特模式，如下表所示：

信源码元	码字
Z_0	00...000
Z_1	00...001
Z_2	00...010
...	⋮
Z_{n-1}	11...111

在这种情况下，编码使用 $M/\text{码元}$ 。显然一个唯一的表示需要 $M \geq \log_2(N)$ 。

更好的编码是给每个信源码元分配长度可变的码字。较短的码字被分配给概率高的码元；较长的码字被分配给概率较低所码元。举一个例子，假设一个信源具有字母表 $Z=\{a,b,c,d\}$ 和概率 $p_a=1/2$ ， $p_b=p_c=p_d=1/6$ 。用于该信源的一个可能的长度可变编码将是：

信源码元	码字
A	0
B	10
C	110
D	111

对于长消息，预期编码长度 L 由 $L=\sum p_n l_n$ 给出，以比特/信源码元，在此 l_n 是编码码元 z_n 的长度。相对于将需要 2 比特/码元的直接二进制编码的编码长度，这样更好。

在上述例子中，使用公知的哈夫曼算法生成码字。最终的码字分配被称为用于该信源的哈夫曼编码。在最小化在所有可能的长度可变

码字中预期编码长度 L 的意义上，哈夫曼编码是最佳的。熵是信源内部信息内容的测量值。通过 $E = -\sum p_n \log_2(p_n)$ 以比特每码元来测量熵。编码理论规定用于任一编码的预期编码长度不能少于信源熵。例如上述信源，熵是 $E = -(1/2)\log_2(1/2) - (1/2)\log_2(1/6) = 1.793$ 比特/码元。可以看出哈夫曼编码长度非常接近于最佳。

另一种可能编码是将固定长度的码字分配给信源码元串。这种串具有可变长度，并且编码效率得自频繁出现的长串可以仅用一个码字来替代。一个例子是下表中的编码。对于该编码，码字始终具有四个比特，但它被用于表示不同长度的串。根据上表中的概率能够轻易地计算出平均信源串长度，将是 $K=25/12=2.083$ 。因为这些串用四个比特表示，比特速率是 $4*12/25=1.92$ 比特/码元。

信源串	串概率	码字
D	1/6	0000
Ab	1/12	0001
Ac	1/12	0010
Ad	1/12	0011
Ba	1/12	0100
Bb	1/36	0101
Bc	1/36	0110
Bd	1/36	0111
Ca	1/12	1000
Cb	1/36	1001
Cc	1/36	1010
Cd	1/36	1011
Aaa	1/8	1100
Aab	1/24	1101
Aac	1/24	1110
Aad	1/24	1111

在上述例子中，使用 B.P.Tunstall 的标题为“无噪声压缩编码合

成”(博士论文, 乔治亚州大学, Technol, 乔治亚州亚特兰大市, 1967年) 参考文献中的技术确定用每个码字(即字符串表)映射字符串的选择。使用该表的编码称作 Tunstall 编码。在最小化在所有可能的长度可变到固定编码中的预期编码长度 L 的意义上, 可以看出 Tunstall 编码是最佳的, Tunstall 编码可以被视为双哈夫曼编码。

在这个例子中, Tunstall 编码可能没有哈夫曼编码有效, 然而, 可以表明 Tunstall 编码的性能随着编码长度的增加即随着字符串表长度的增加将接近信源的熵。根据本发明, Tunstall 编码具有相对于哈夫曼编码的优点即更快的解码。这是因为每个码字始终具有相同的比特数, 因此更容易分析(在下文详细讨论)。

因此, 如图 15 所示, 本发明最好使用熵编码器, 它可以是行程编码器和 Tunstall 编码器。即, 图 15 是表示根据本发明用于执行熵编码的系统和方法的流程图。参考图 15 和图 13, 根据本发明, 图 15 表示最好是可变长度熵编码器的一个编码器。

熵是由诸如概率模型的模型提供的信息表示(换句话说, 在消息中所包含信息的测量值)。最佳熵编码器生成在一个消息中用一个码元表示的平均信息量, 并且是用于生成消息的概率模型的函数(在下文更详细地说明)。增加模型的复杂性以便模型更好地反映信源码元在始发消息中的实际分配以减少消息。最佳熵编码器利用行程编码器继之以诸如常规 Tunstall 编码器的可变-固定长度编码器编码量化系数。

行程编码器为零序列降低码元速率。可变-固定长度编码器从信源输出的可变长度字符串的一个字典中映射到一组给定长度的码字。可变-给定长度编码使用与信源输出相关的统计。Tunstall 编码器使用可变-固定长度编码器为离散、无后效的信源最大化每字典字符串的信源字母预期数。换句话说, 输入序列被划分长可变长度块以最大化平均消息长度, 并将每块分配给固定长度编码。

先前的编码器, 例如 ASPEC, 使用对变换系数子集的行程编码, 并用诸如哈夫曼编码器的矢量固定-可变长度编码器编码非零系数。

相反, 本发明最好使用行程编码器, 它对由所有量化变换系数构成的矢量执行操作, 实质上是建立一个新码元信源, 其中量化零值的游程被用定义游程长度的码元替代。当游程中零的数量在范围 $[R_{\min}, R_{\max}]$ 中时, 本发明的行程编码器用特定码元替代零的游程。某种情况下, 例如通过简单地设置 $R_{\max} < R_{\min}$ 可以关闭行程编码器。

Tunstall 编码器没有被广泛地使用, 因为编码器的效率与信源码元的概率模型直接相关。例如, 当设计用于压缩的编码时, 如果对于信源有好的模型可以进行更有效的编码, 即模型越好, 压缩越好。结果, 为了有效地编码, 需要好的概率分布模型以为编码器建立一个合适的字符串字典。如下所述, 本发明使用一个充分概率模型, 它使 Tunstall 编码切实可行和有效。

一般来说, 如上所述, 量化系数被用行程编码器继之以可变-固定长度块编码器进行编码。具体地说, 首先, 由计算模块作为一个数据块接收量化变换系数 $q(k)$, 用于为该块计算最大绝对值 (方框 1510)。即, 扫描所有量化值以确定最大幅值 $A = \max |X_r(k)|$ 。然后, 由近似模块量化 A (方框 1512), 该近似模块用于使用 $vr \geq A$ 来近似 A , vr 是 $[4, 512]$ 范围内的 2 的幂。从而用 3 个比特编码 vr 的值并发送给解码器。接着, 替换模块接收 $q(k)$ 并连接到近似模块, 用表示游程长度的可变-固定长度编码字典中定义 (图 16 的方框 1610, 将在下文详细描述) 的新码元替代范围 $[R_{\min}, R_{\max}]$ 中零的游程 (方框 1514)。利用根据本发明的参数建模技术计算这个字典, 如下文在图 16 中描述中。随后, 用于诸如 Tunstall 编码器的可变-固定长度编码器编码结果值 $s(k)$ (方框 1516), 以生成信道码元 (信息比特)。另外, 因为熵编码器的效率直接取决于所用的概率模型, 需要结合根据本发明的好的参数模型, 如下文将详细讨论的。

参数建模

图 16 是表示根据本发明用于使用概率建模执行熵编码的系统和方法的流程图。如上所述, 熵编码器的效率与概率模型的质量直接相

关。如图 16 所示，编码器需要一个输入字符串的字典，它可以用简单算法来建立以根据码元概率编辑一个输入字符串字典（下文详细说明）。尽管可以使用算术编码器或哈夫曼编码器，诸如上述 Tunstall 编码器的可变-固定长度编码器可以使用本发明的参数模型和简化解码实现接近算术编码器的效率。这是因为所有的 Tunstall 码字具有相同的长度，例如它可以被设置为一个字节。

而且，与诸如清晰话音的简单信号相比，当前的变换编码器通常对诸如音乐的复杂信号可以更加有效地执行。这是因为与这种信号相关的更高的屏蔽级别和由当前变换编码器使用的熵编码类型。因此，对于清晰话音，以低比特率工作的当前变换编码器可能不能再生优良的谐波结构。即对于有声话音和以大约 1 比特/抽样的速率，量化步长足够大以使大多数变换系数量化为零，除了基本声域频率的谐波之外。然而，使用上述熵编码器和下述的参数建模，与诸如一阶编码器的当前熵编码系统所推算的结果相比，本发明能够生成更好的结果。

一般来说，本发明的参数建模使用量化和游程长度编码变换参数的概率分布函数（PDF）的模型。通常，使用熵编码（一般是哈夫曼编码）的编解码器从音频抽样集中获取的频率分布图中得出 PDF（和它们相应的量化表）。相反，本发明使用适合每个输入块的改进拉普拉斯算子 + 指数概率密度，它允许更好的编码性能。本发明的 PDF 模型的一个优点是它的形状由单个参数控制，该参数与量化系数的峰值直接相关。这导致不需要模型选择的计算开销，并且实际上不需要给解码器指定模型的开销。最后，本发明使用二进制搜索程序确定最佳量化步长。如下所述的二进制搜索程序比先前的方法简单得多，先前的方法例如在每次迭代中执行与屏蔽阈值相关的附加计算的方法。

具体地说，本发明的概率分布模型最好使用一个改进拉普拉斯算子 + 指数概率密度函数（PDF）以适合每个输入块的量化变换系数的频率分布图。PDF 模型由在上述图 15 的方框 1510 中所述的参数 A 控制（注意 A 用 v_r 近似，如图 15 的方框 1512 所示）。因此，PDF 模型由下式定义：

$$\Pr(s=m)=\begin{cases} \beta_1 \left[\exp(-d_L(m-A)^{0.9}-1) \right] + 0.01 & m \leq 2A, m \neq A \\ 0.25, m=A \text{ (or } q=0) & \\ \beta_2, 2A+2 \leq m \leq 2A+4 & \\ \beta_2 \left[\exp(-d_R(m-2A-4)^{0.8}-1) \right] + 0.01 & m \geq 2A+4 \end{cases}$$

其中变换和游程编码码元 s 属于下述字母表:

量化值	码元
$-A, -A+1, \dots, A$	$0, 1, \dots, 2A$
$R_{\min}$ 个零的游程	$2A+1$
$R_{\min}+1$ 个零的游程	$2A+2$
....	
$R_{\max}$ 个零的游程	$2A+1+R_{\max}-R_{\min}$

对于用于步长优化的二进制搜索，如上所述在标量量化中使用的量化步长 dt 控制创建保真度和比特速率之间的权衡。较小的量化步长产生较好的保真度和较高的比特速率。对于固定速率应用，需要重复调整量化步长直到码元编码器 (Tunstall) 输出上的比特速率尽可能地匹配希望的速率 (而不超过它)。

几种技术可用于调整步长。一种技术包括：1) 开始于量化步长，用 dB 表示， $dt=dt_0$ ，其中 dt_0 是根据输入定标的一个参数。2) 设置 $kdd=16$ ，并检查用 dt 获得的速率。如果超过预算，通过 $dt=dt+kdd$ 改变步长，或者通过 $dt=dt-kdd$ 改变它。3) 重复上述步骤，在每次迭代时将 kdd 除以 2 (二进制搜索)，直到 $kdd=1$ ，即以 $1dB$ 内的精度确定最佳步长。很容易看出这个程序可以生成最多 64 个不同步长，因此，最佳步长用 7 个比特表示并发送给解码器。

再次参考图 6，表示根据本发明用于解码音频信号的系统的总方框/流程图。该解码器使用合适的相反处理步骤，如图 6 所示。可变-固定长度解码器 (例如 Tunstall 解码器) 和游程解码模块接收编码比特流和与 PDF 范围参数相关的辅助信息，用于恢复量化变换系数。连接到可变-固定游程长度解码器的均匀去量化模块和游程长度解码模块从均匀量化重建，用于恢复加权 NMLBT 变换系数的近似值。一个反加权模块执行反加权，用于使变换系数返回它们用于反变换的合

适的标度范围。反 NMLBT 变换模块将近似值恢复到初始信号块。可用信道比特率越大，量化步长越小，因而重建的保真度越好。

应当注意到因为两个原因使解码器的计算复杂性比编码器低。第一，诸如 Tunstall 解码的可变-固定长度解码（仅需要表查找）比它的对应编码（需要串查找）快。第二，因为已知步长，仅使用一次去量化（不需要循环，与编码器不同）。然而，无论如何，对于编码器和解码器，在 NMLBT 中有大量的计算，它可以通过快速付立叶变换来有效地计算。

已经为说明和描述目的进行了本发明的上述说明。它将不是穷尽的，也不是将本发明限制到所公开的具体形式。在上述教导的启示下可以进行多种修改和变化。本发明的范围将由权利要求书而非详细说明来限制。

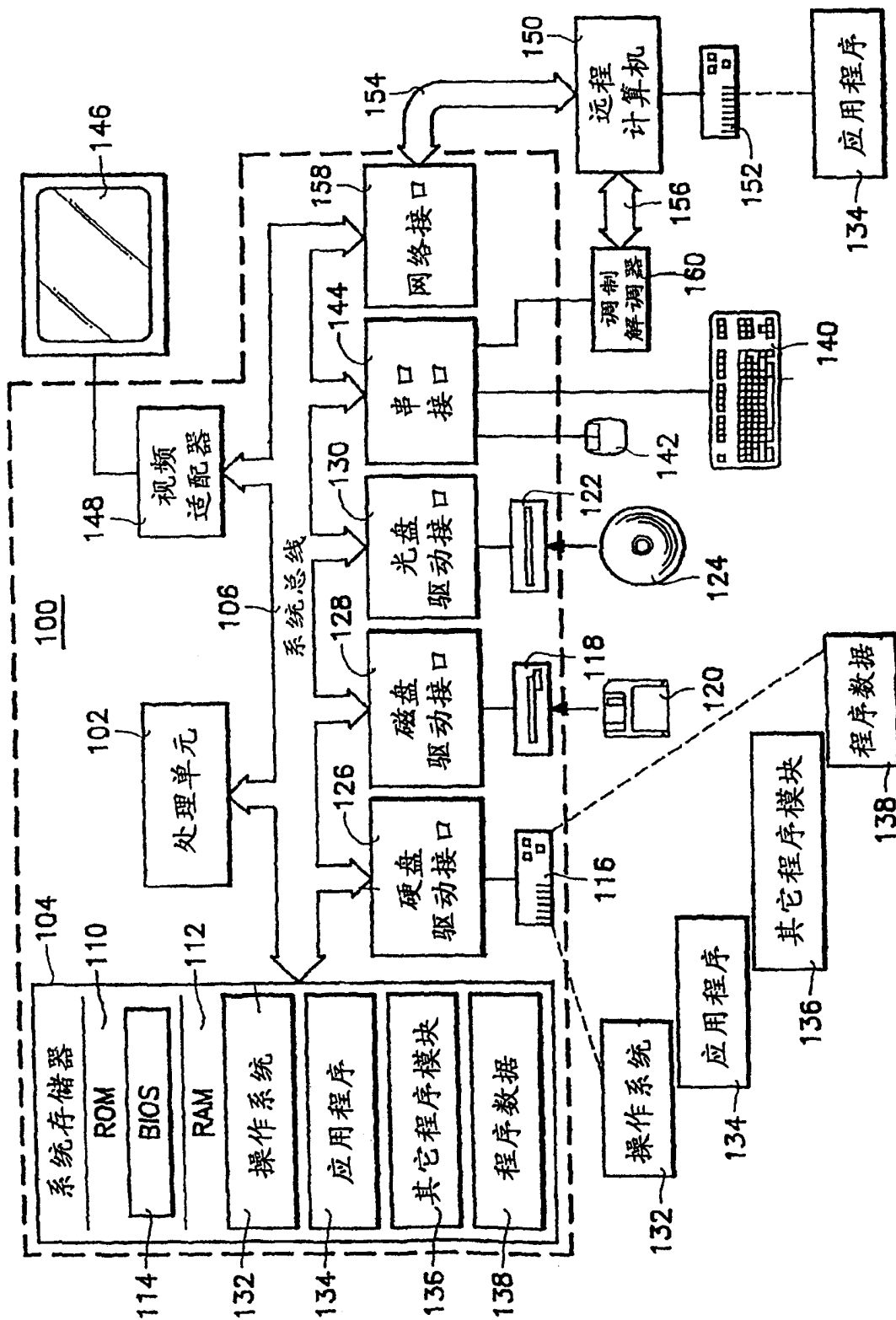
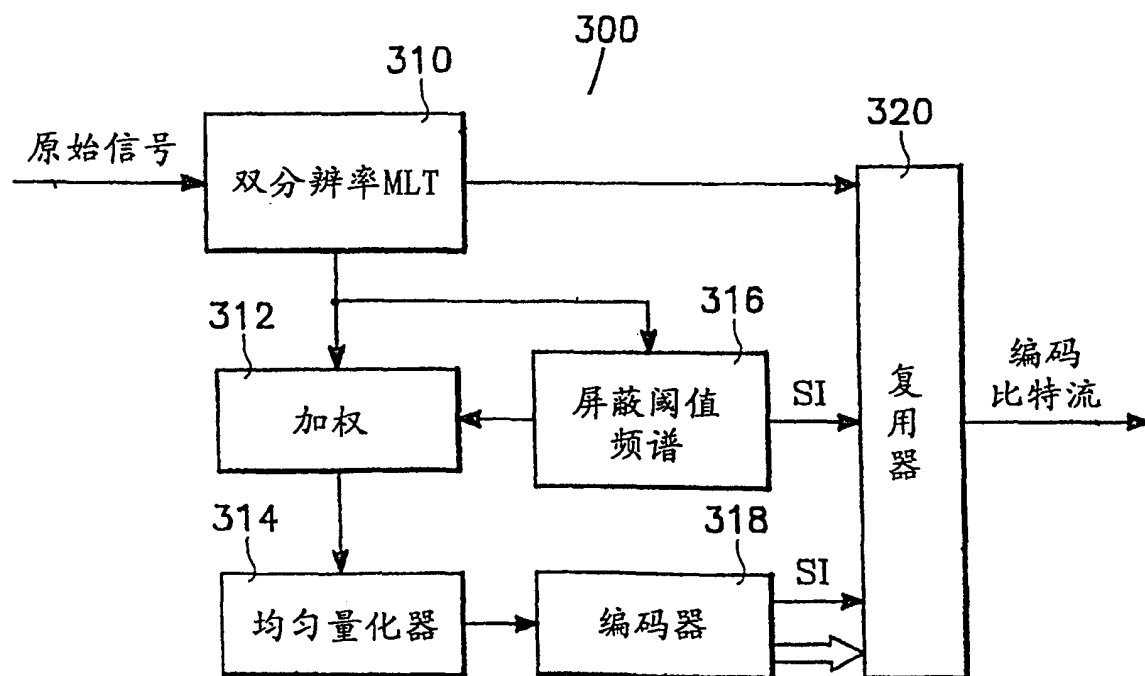
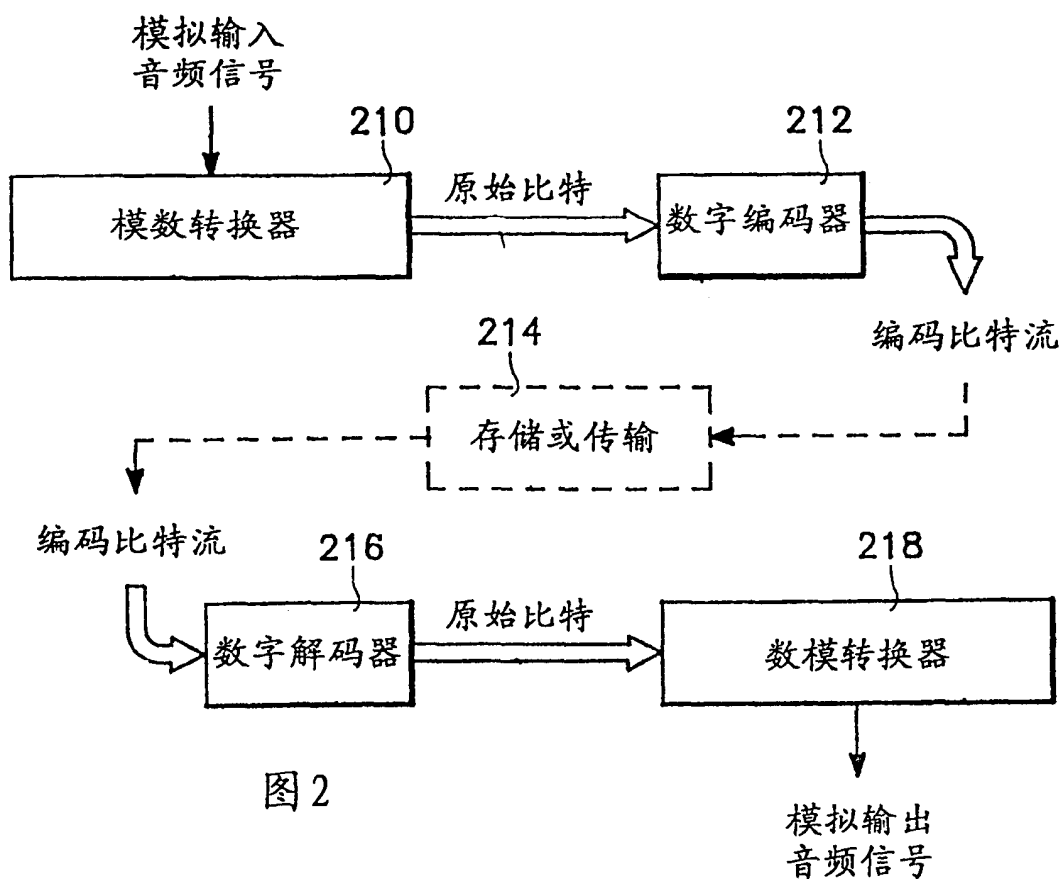


图1



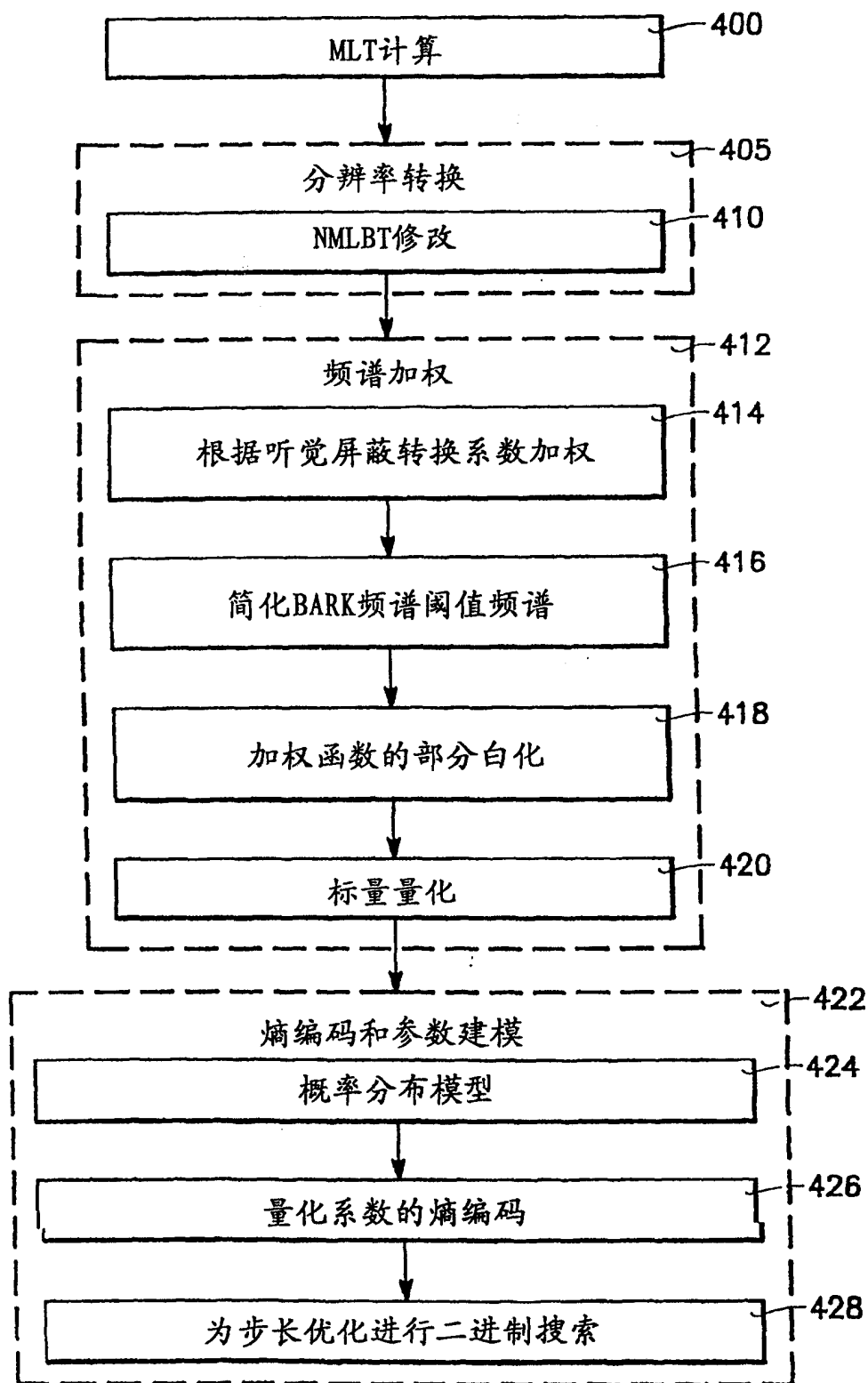


图 4

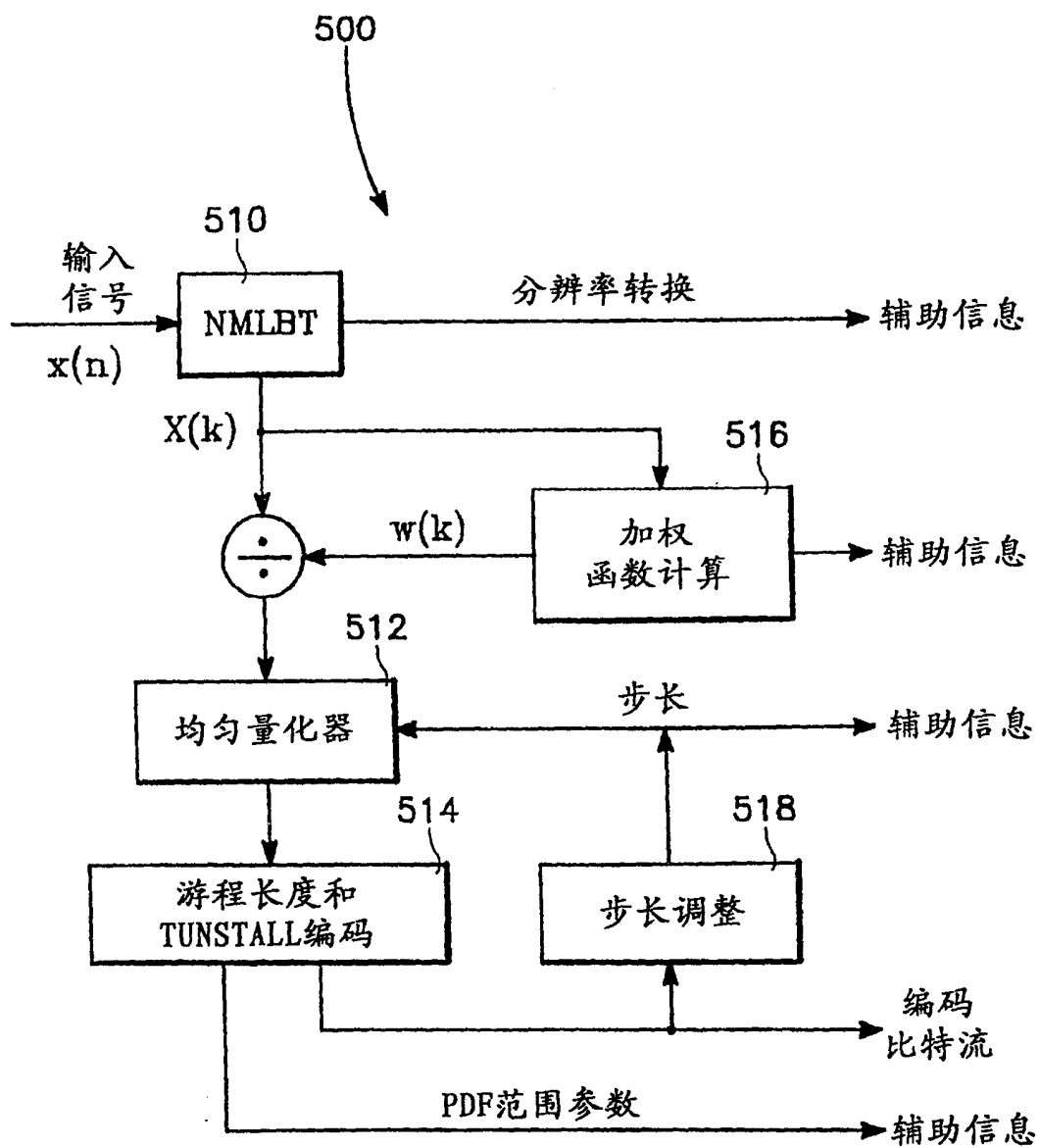


图5

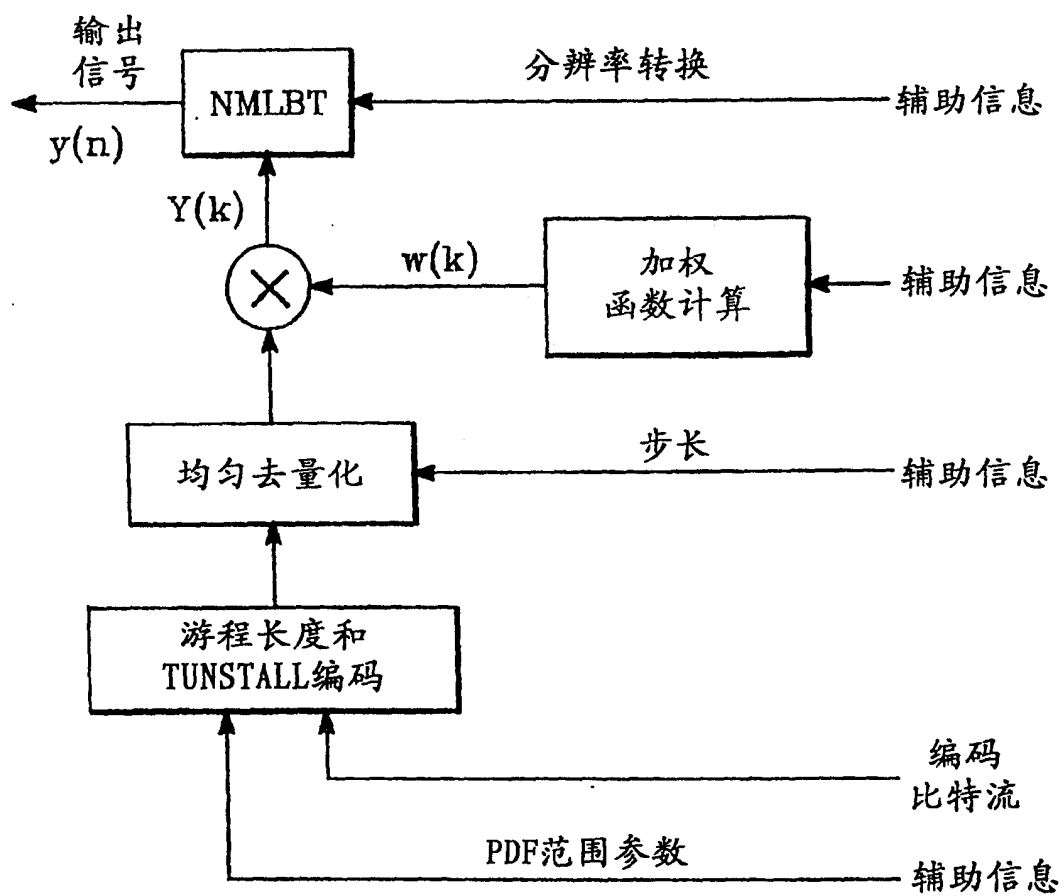


图6

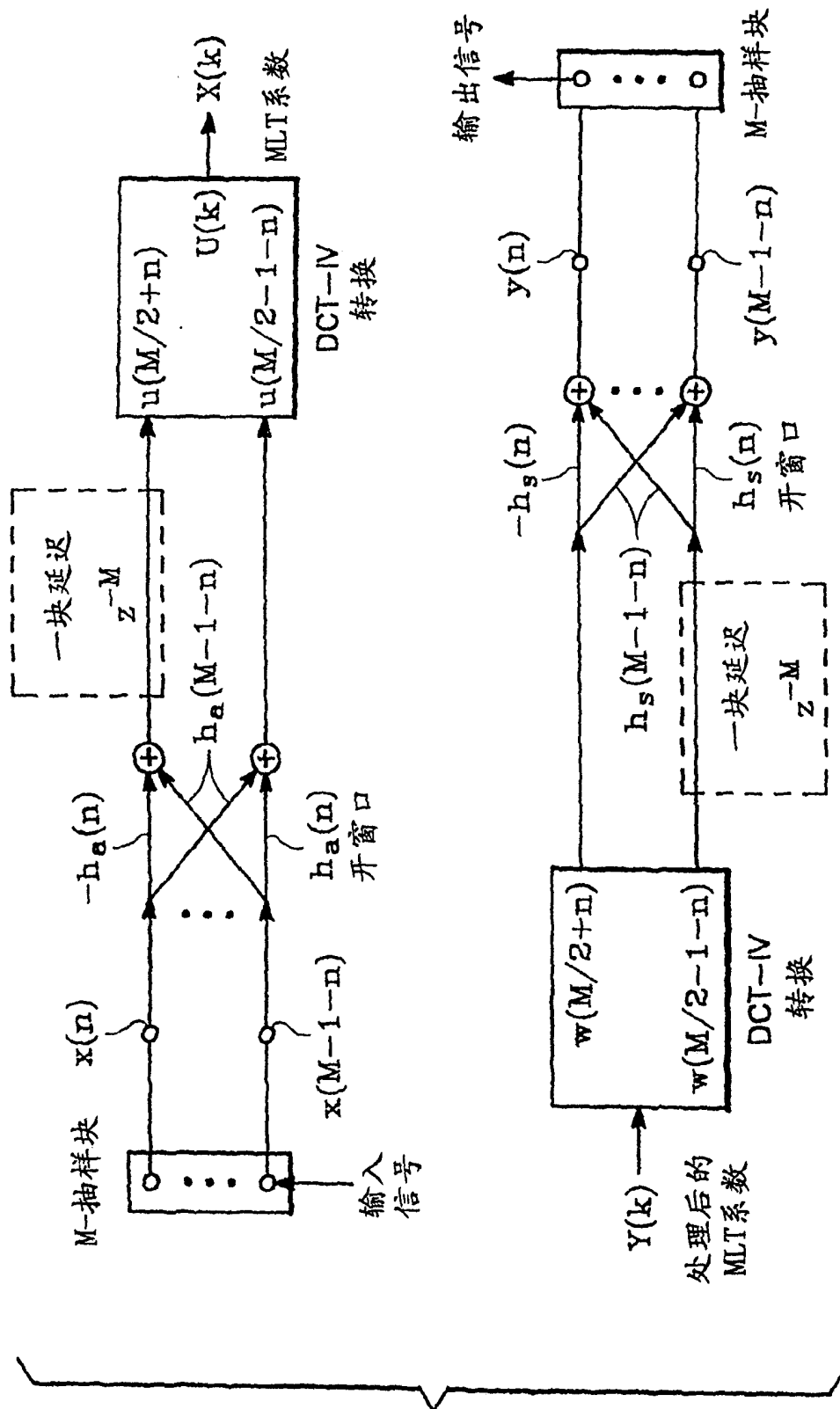


图7

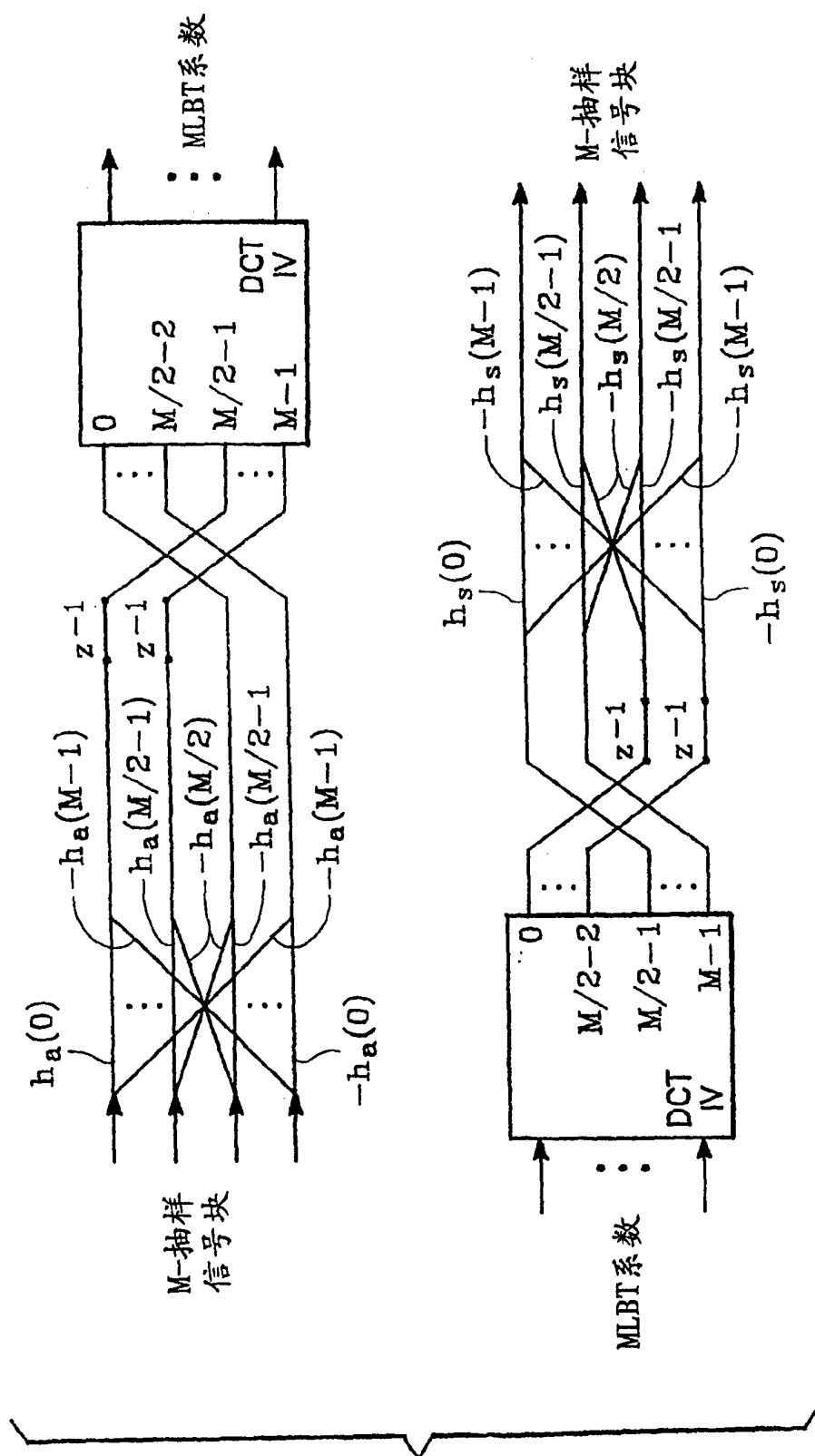


图 8

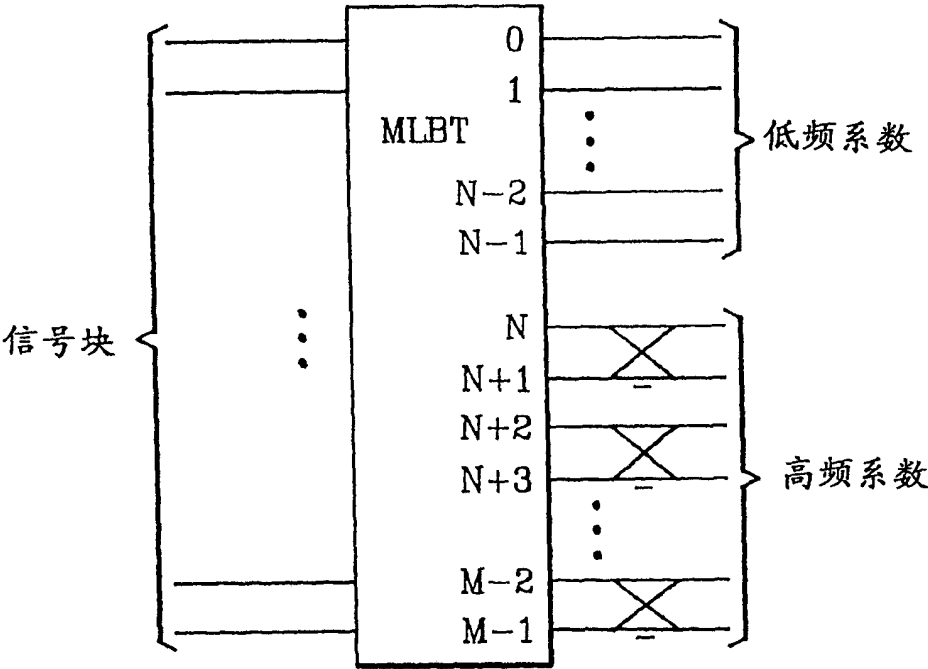


图 9

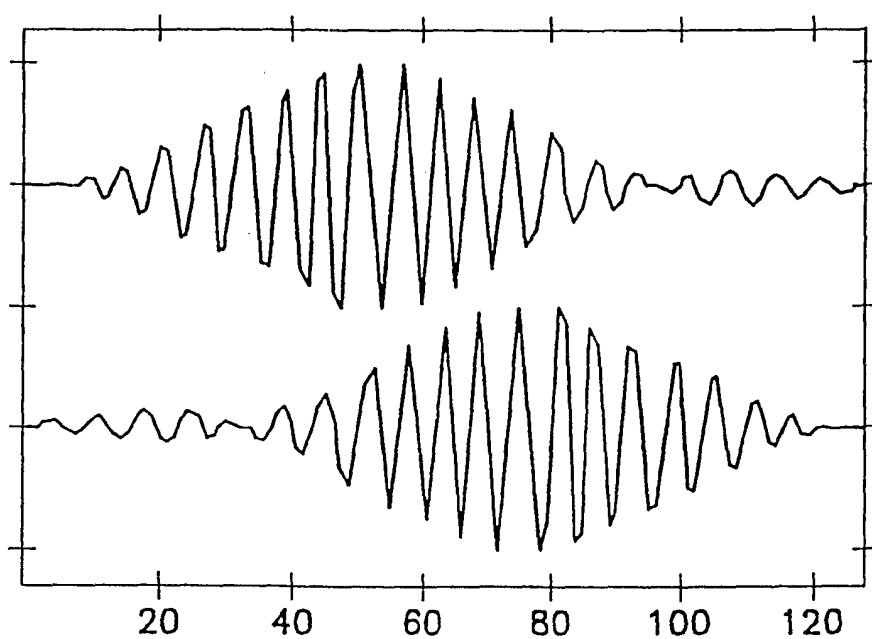


图 10

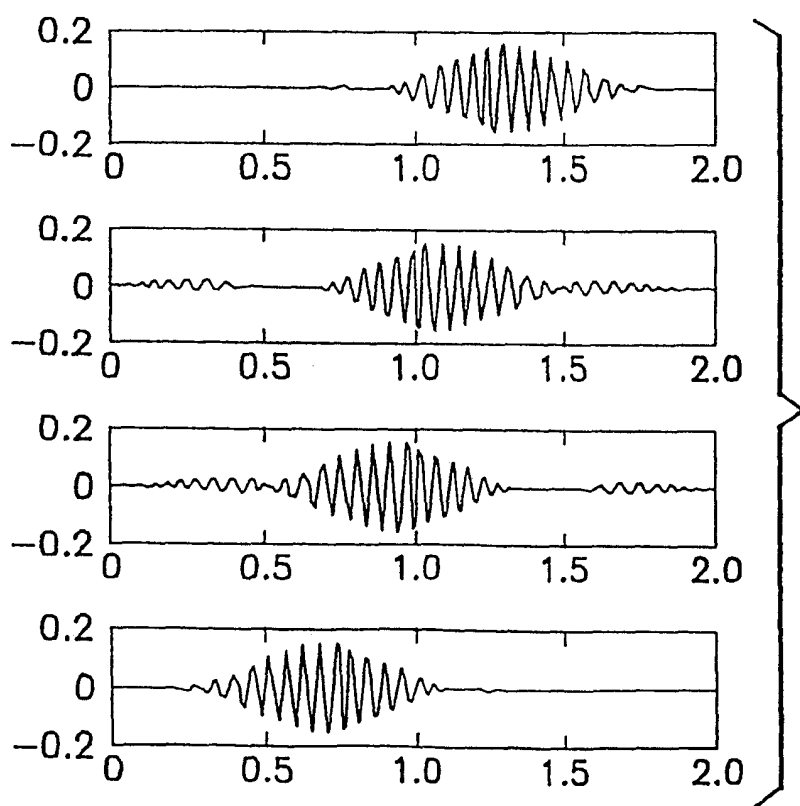


图 11

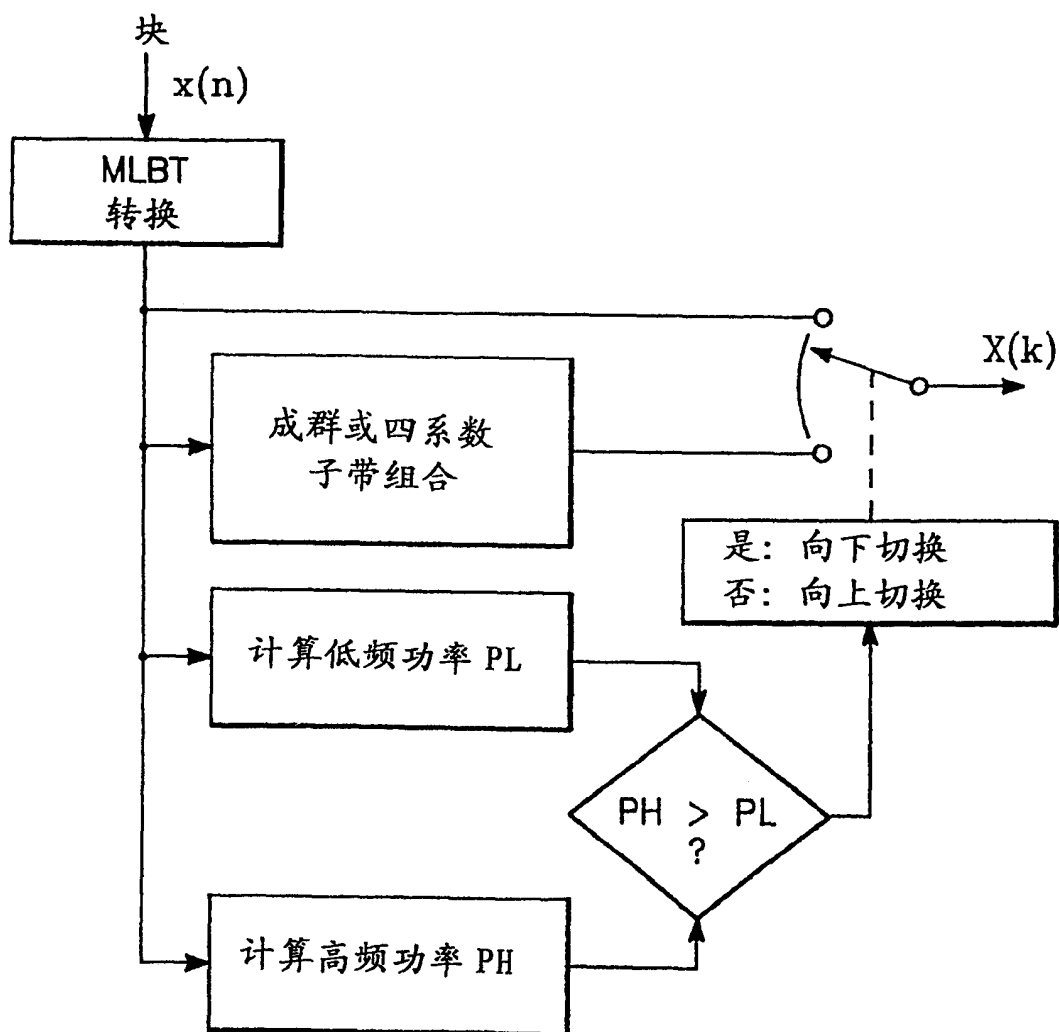


图 12

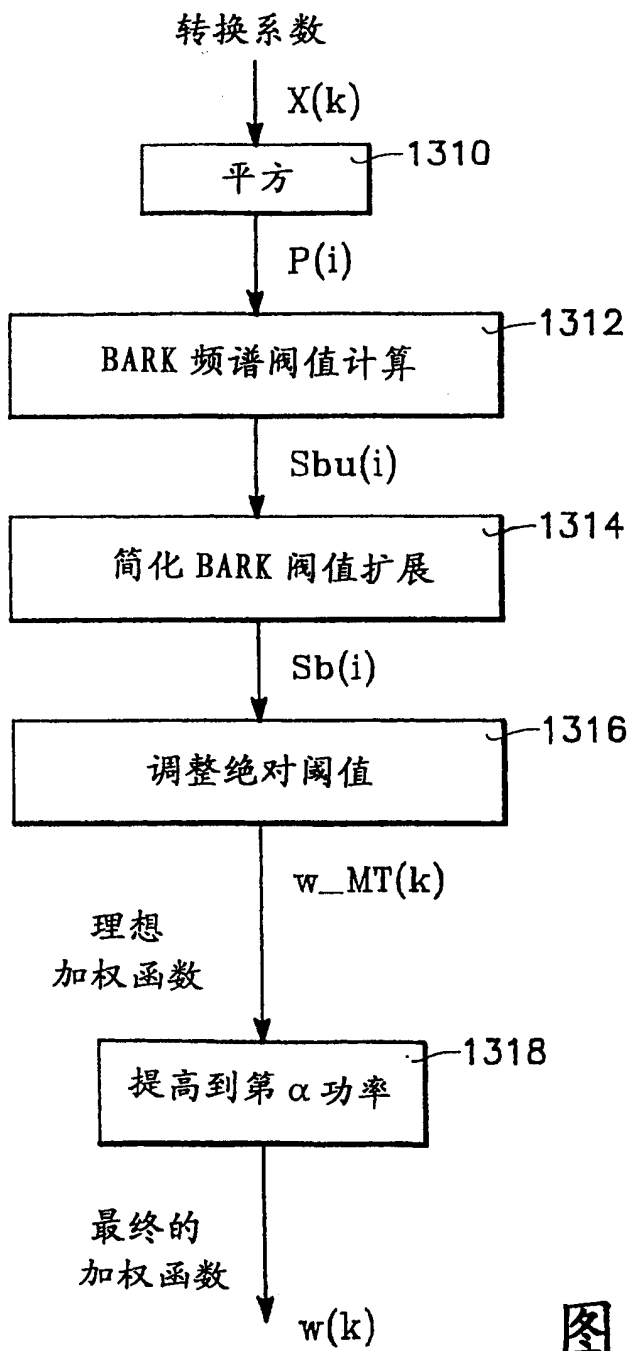


图 13

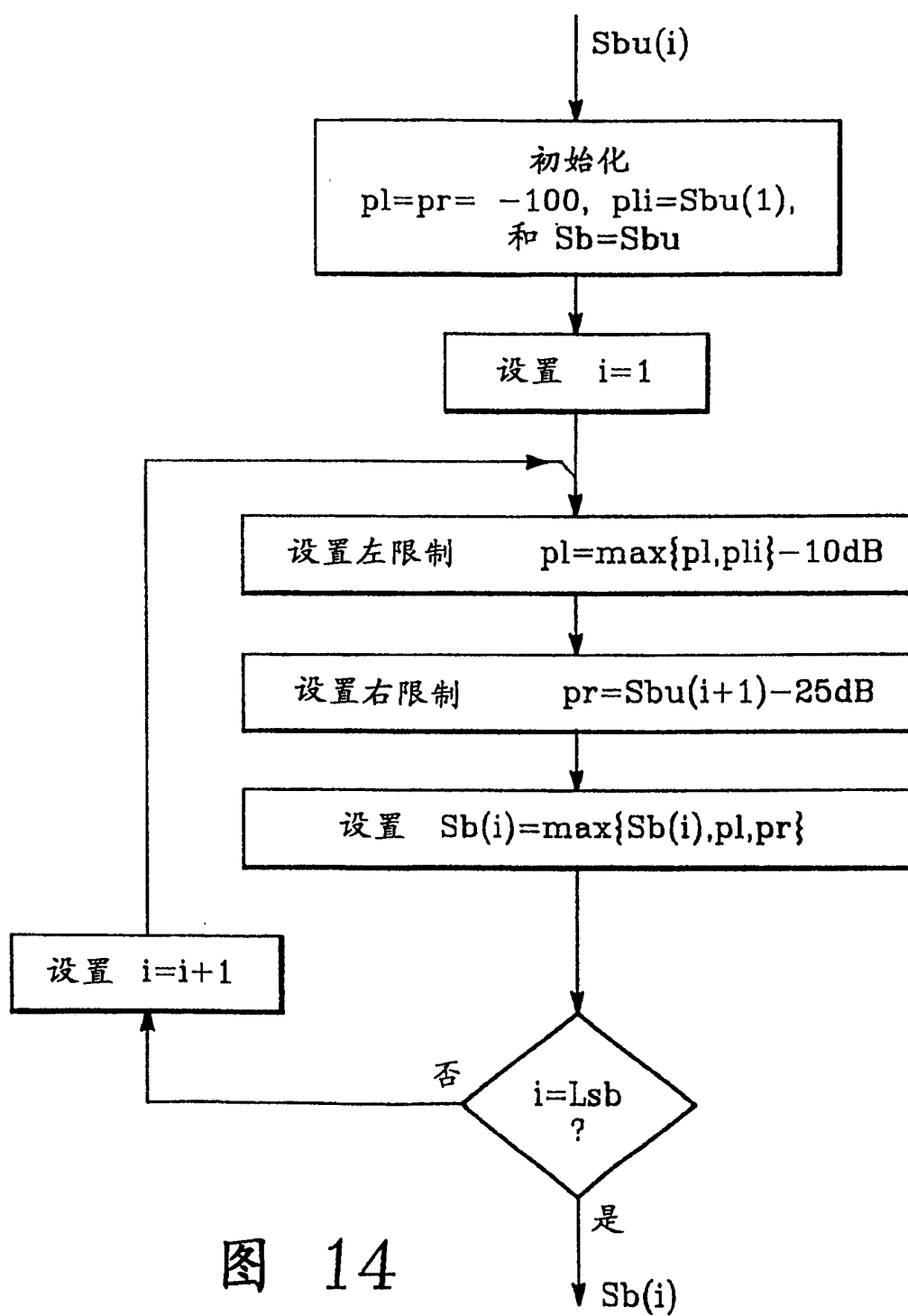


图 14

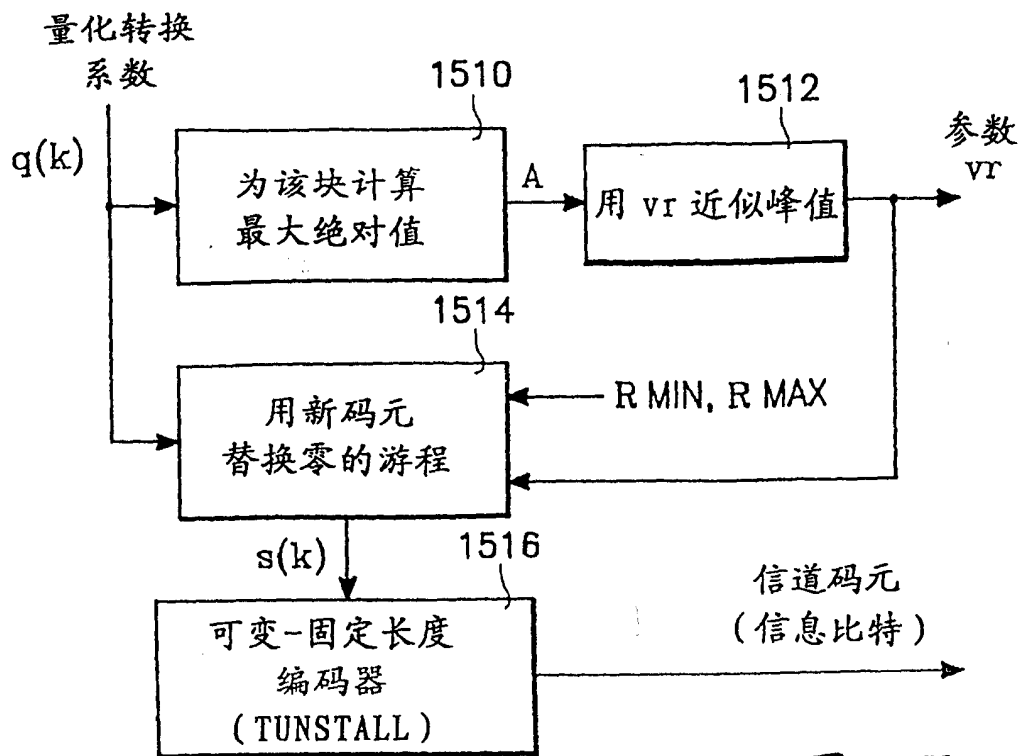


图 15

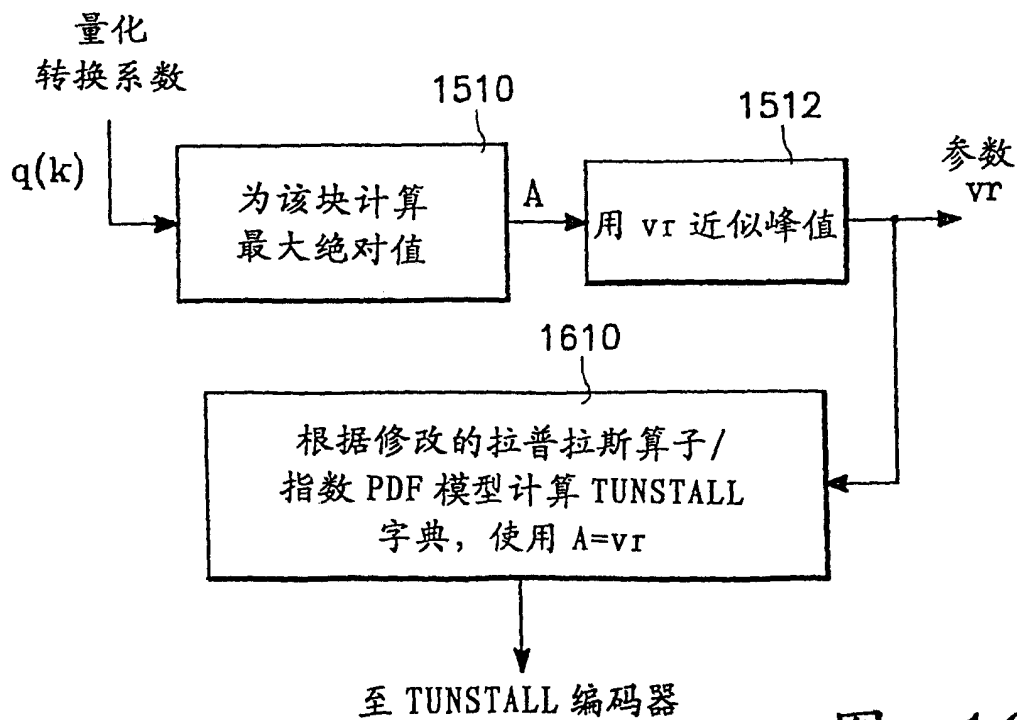


图 16