



[12] 发明专利申请公开说明书

[21] 申请号 02828522.0

[43] 公开日 2005 年 6 月 1 日

[11] 公开号 CN 1623185A

[22] 申请日 2002.3.12 [21] 申请号 02828522.0

[86] 国际申请 PCT/IB2002/000723 2002.3.12

[87] 国际公布 WO2003/077235 英 2003.9.18

[85] 进入国家阶段日期 2004.9.10

[71] 申请人 诺基亚有限公司

地址 芬兰埃斯波

[72] 发明人 S·施特赖希 M·维莱尔莫

[74] 专利代理机构 中国专利代理(香港)有限公司

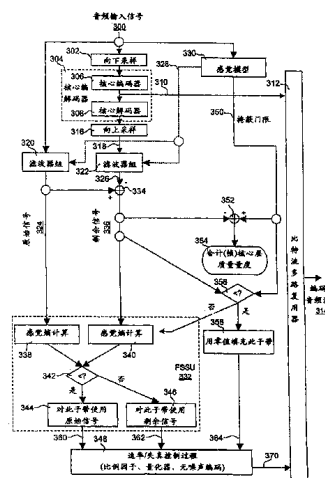
代理人 杨凯 张志醒

权利要求书 5 页 说明书 15 页 附图 6 页

[54] 发明名称 可伸缩音频编码的有效改进

[57] 摘要

一种将音频信号编码成具有第一层和第二层的分层数据流的音频编码方法。第二层用于增强第一层。该方法包括：形成原始数字音频信号(300)；对原始数字音频信号编码以获得第一层信号(310)；生成剩余信号(336)以反映原始信号与第一层信号之间的差值；选择原始信号或剩余信号来编码；以及通过对所选信号编码而得到第二层信号。此外，对剩余信号(336)进行评估，如果评估结果满足给定判据，则选择预定的低熵信号以将其编码成第二层信号，而不选择原始信号或剩余信号。



1. 一种将音频信号编码成具有第一层和第二层的分层数据流的音频编码方法，所述第二层用于增强第一层，所述方法包括如下步骤：
- 5 形成原始数字音频信号（300）；
对所述原始信号（300）编码以获得第一层信号（310）；
生成反映所述原始信号与所述第一层信号之差的剩余信号（336）；
- 10 选择所述原始信号或所述剩余信号来编码；以及
通过对所选信号编码而生成第二层信号；
其特征在于：
对所述剩余信号（336）进行评估；以及
如果所述评估结果满足给定判据，则
- 15 选择预定的低熵信号而非所述原始信号或所述剩余信号，以将其编码成所述第二层信号。
2. 如权利要求 1 所述的方法，其特征在于：所述原始数字音频信号（300）和所述剩余信号（336）包括多个频率子带（401-449），并且其中分别对所述多个频率子带执行评估所述剩余信号以及选择
- 20 所述预定低熵信号的所述步骤。
3. 如权利要求 1 或 2 所述的方法，其特征在于，所述评估步骤包括将所述剩余信号（336）与掩蔽门限（350）作比较，其中，所述判据包括发现所述剩余信号低于所述掩蔽门限。
4. 如权利要求 1-3 中任意一项所述的方法，其特征在于，所述
- 25 评估步骤包括计算对所述原始信号（324）或所述剩余信号（336）量化所引起的误差。
5. 如权利要求 4 所述的方法，其特征在于，所述判据包括量化引起的所述误差是可感觉的。

6. 如权利要求 4 或 5 所述的方法，其特征在于，所述判据包括量化引起的所述误差大于所述剩余信号。
7. 如权利要求 2-6 中任意一项所述的方法，其特征在于还包括如下步骤：
- 5 提高所述多个频率子带（401-449）中至少一个频率子带的比特率，对于该频率子带，所述剩余信号未被所述预定一致信号替代。
8. 如权利要求 2-6 中任意一项所述的方法，其特征在于还包括如下步骤：降低所述多个频率子带（401-449）整体的比特率。
9. 如权利要求 2-6 中任意一项所述的方法，其特征在于还包括如下步骤：减少所述多个频率子带（401-449）中至少一个频率子带的量化误差，对于该频率子带，所述剩余信号未被所述预定一致信号替代。
- 10 如下步骤：减少所述多个频率子带（401-449）中至少一个频率子带的量化误差，对于该频率子带，所述剩余信号未被所述预定一致信号替代。
10. 如前述任意一项权利要求所述的方法，其特征在于，所述预定低熵信号是恒定低振幅信号。
- 15 11. 如权利要求 10 所述的方法，其特征在于，所述预定低熵信号具有零值振幅。
12. 如前述任意一项权利要求所述的方法，其特征在于，所述第一层信号是核心层信号。
13. 如权利要求 12 所述的方法，其特征在于，所述第一层信号是通过自适应多速率宽带（AMR-MB）编码得到的。
- 20 14. 如前述任意一项权利要求所述的方法，其特征在于，所述第二层信号是通过 AAC（高级音频编码）编码得到的。
15. 如权利要求 2-14 中任意一项所述的方法，其特征在于还包括如下步骤：
- 25 对所述多个频率子带（401-449），通过将所述各频率子带的所述掩蔽门限（350）和所述剩余信号（336）之间的差值相加来累积核心层质量量度（710）；以及
- 在评估所述剩余信号（336）时使用所述核心层质量量度。

16. 如前述权利要求中任意一项所述的方法，其特征在于还包括如下步骤：将所述第一层信号解码以获得解码的第一层信号（310），其中，所述剩余信号（336）反映所述原始信号和所述解码的第一层信号之间的差值。

5 17. 一种可直接装入处理器（240）的存储器（242）中的计算机程序产品，其中，所述计算机程序产品包括在由所述处理器执行时用于执行如权利要求 1-16 中任意一项所述的方法的程序代码。

18. 一种集成电路，该集成电路适于执行如权利要求 1-16 中任意一项所述的方法。

10 19. 一种音频编码器，用于将音频信号编码成具有第一层和第二层的分层数据流，所述第二层用于增强所述第一层，所述编码器包括：

可以将原始数字音频信号（300）编码成第一层信号（310）的第一编码单元（304）；

15 用于生成反映所述原始信号与所述第一层信号之差的剩余信号（336）的部件（334）；

适于选择所述原始信号或所述剩余信号来编码以将其编码成第二层信号的第一选择器（332）；以及

20 可以通过对所选信号编码而生成所述第二层信号的第二编码部件；

其特征在于：

用于评估所述剩余信号（336）并提供所述评估结果满足给定判据的指示的部件；以及

25 第二选择器（356），所述第二选择器与所述评估部件相关联，且适于在收到所述指示时选择预定低熵信号而非所述原始信号或所述剩余信号，以由所述第二编码部件加以编码。

20. 如权利要求 19 所述的音频编码器，其特征在于：所述原始数字音频信号（300）和所述剩余信号（336）包括多个频率子带

(401-449)，其中，所述用于评估所述剩余信号的部件和所述第二选择器(356)适于对所述多个频率子带中的各个频率子带起作用。

21. 如权利要求 19 或 20 所述的音频编码器，其特征在于，所述第一编码部件是自适应多速率宽带 (AMR-WB) 编码器。

5 22. 如权利要求 19-21 中任意一项所述的音频编码器，其特征在于，所述第二编码部件是 AAC (高级音频编码) 编码器

23. 一种包括音频解码器的音频码变换器，其中，所述音频解码器适于接收音频输入信号并将其解码，该音频输入信号是根据本质上已知的音频编码技术来编码的，其特征在于：

10 如权利要求 19-22 中任意一项所述的音频编码器。

24. 一种音频解码器，包括：用于接收包含编码的第一层信号和至少一个编码的第二层信号的分层数据流的部件、可以将所述第一层信号解码的第一解码部件以及可以将所述第二层信号解码的第二解码部件，其特征在于：

15 用于确定所述第二层信号的感觉质量量度的部件；

 用于判断所述感觉质量量度是否满足预定判据的部件；

 用于在所述感觉质量量度满足所述判据时将所述第二层信号与所述第一层信号合并以得到解码输出信号的部件；以及

20 用于在所述感觉质量量度不满足所述判据时只根据所述第一层信号生成所述解码输出信号的部件。

25. 如权利要求 24 所述的音频解码器，其特征在于，所述感觉质量量度包含在所述接收的分层数据流中。

26. 如权利要求 24 所述的音频解码器，其特征在于还包括：用于通过识别所述解码的第二层信号中的非自然音频分量而导出感觉质量量度的部件。

27. 如权利要求 26 所述的音频解码器，其特征在于，所述非自然音频分量包含毛刺、噪声或非预期波形中的至少一项。

28. 一种用于移动电信网 (110) 的站 (200)，其包括根据权

利要求 19-22 中任一权利要求的音频编码器、根据权利要求 24 的音频解码器和根据权利要求 23 的音频码变换器中的至少一项。

29. 如权利要求 28 所述的站, 其特征在于, 所述站是基站(104)。

30. 如权利要求 28 所述的站, 其特征在于, 所述站是移动终端
5 (100, 112)。

可伸缩音频编码的有效改进

5 发明领域

本发明涉及一种将音频信号编码成具有第一层和第二层的分层数据流的音频编码方法，其中，所述第二层用于增强所述第一层。更具体地说，本发明涉及一种音频编码方法，其中，将原始数字音频信号编码以获得第一层信号，生成反映原始信号和第一层信号之差的剩余信号，并选择原始信号或剩余信号以便将其编码成第二层信号。

背景技术

音频即声学能量本质上是模拟的。但是，将音频表示成数字形式对于存储或传输均很方便。通过对模拟音频信号采样和数字化而获得的纯数字音频数据需要很大的存储空间和信道带宽，对高质量音频尤其如此，高质量音频例如可以按采样频率为 44 kHz、每样本 16 比特（普通的音频 CD 质量）的形式来表示。因此，数字音频通常是按各种已知的信源编码技术来压缩的。

感觉音频编码技术、如 MPEG 层-3 (MP3)、MPEG-2 和 MPEG-4 均利用了人耳的信号掩蔽特性来减少数据量。这样，就将量化噪声分散到各频率子带，从而被总信号所掩蔽，即噪声仍然是不可闻的。可以在几乎没有或几乎察觉不到音频质量下降的情况下大大减少存储容量。

感觉音频编码技术通常是可伸缩的，可生成具有基础层和至少一个增强层的分层比特流。这可实现比特率伸缩，即在解码器侧以不同的音频质量等级解码或通过流量整形或调节而降低网络中的比特率。一种方法是提供纯单声道的基础层编码并提供在音频中增加

立体声质量的增强层编码。这样，可以在解码器侧选择仅对基础层信息解码（例如，如果解码器侧的接收装置只有一个扬声器），或者选择同时对基础层信息以及增强层信息解码以生成立体声声音。

在可伸缩音频编码环境中，“基础层”和“核心层”是同义词。

- 5 ISO/IEC 14496-3:2001(E)的子部分 4 中描述了一部分 MPEG-4 音频标准，并提出了一种 MPEG-4 兼容的核心编解码器或 CELP（码激励线性预测）类型的外部核心编解码器与 AAC（高级音频编码）增强层编解码器的结合方式，以提供高效的比特率可伸缩性。

10 AMR-MB（自适应多速率宽带）语音编解码器是 CELP 型的编解码器的一个实例，该编解码器将用于第三代移动终端中，在第三代合作计划（3GPP）TS 26.190 V5.0.0（2001-03）中对此有所描述。

15 在所述 MPEG-4 音频标准中提及的可伸缩音频编码装置中，增强层编码器中的频率选择切换单元（FSSU）估计对原音频信号或剩余信号编码所需的比特量，所述剩余信号通过将原始信号从前一层（核心层）的重建输出信号中减去而导出。FSSU 总是选择需要较少比特进行编码的备选者。此决定分别针对各个频率子带（即针对表示信号的各固定谱线组）作出。为了允许解码器侧的重建，编码器必须发送 FSS 控制信息，以指示对应于每个音频帧中的每个子带选中了两个备选者中的哪一个。根据此控制信息，随后将来自增强层解码器的输出信号加到仅在剩余信号已被编码的那些子带中的核心层解码器的输出上。

20

25 然而，本发明人已认识到类似于上述的可伸缩音频编码装置的如下问题。具体而言，对于低比特率和中等比特率（例如在 12 kbps-24 kbps 的范围内），有时没有足够多的比特可供以一种使量化误差保持不可察觉的方式来对增强信号编码。在解码器侧，装置误差听起来像啞啞声、爆裂声等，因而非常令人讨厌。实际上，与单独的核心层输出信号相比，所述误差甚至导致可察觉质量的下降。

在现有技术中，为了防止这种效应，或者必须冒损失可闻信息

的风险来限制编码频率范围，或者必须提高增强层编解码器的比特率，提高比特率可能不是所期望的，考虑到可用的网络带宽，甚至可能是不可行的。

5 发明内容

鉴于以上论述，本发明的目的就是解决上述问题或至少减轻上述问题，并以与现有技术同样或低得多的比特率提供改善的声音质量。

一般而言，以上目的是通过如所附独立专利权利要求书所述的一种音频编码方法、音频编码器、音频码变换器、音频解码器、计算机程序产品、集成电路和用于移动通信网的站来实现的。

简单地说，所述目的是这样实现的：除了在现有技术中选择使用剩余信号或原始信号以外，还考虑用于对增强层信号进行编码的其它替代选择。在优选实施例中，所述替代选择包括对于某些频率子带直接将核心层的输出信号作为增强层的输出。这是以如下方式来实现的：用零值或另一类似的低熵信号替换剩余信号，并通过所述 FSS 控制信息指示所述子带的剩余信号已编码。因此，解码器侧不需要任何额外的开销；如果 FSS 控制信息指示剩余信号已编码以及对应的频率子带已在编码器中用零值替换，则对于这些子带，将加上解码器中核心层的输出信号，从而将零值替换。

本发明具有至少两个主要优点：

第一，本发明可用于确保量化引起的误差从不大于只使用前一层所引起的误差。此外，因为量化引起的可闻误差非常令人讨厌，所以当前层的编码器可能被迫仅在某些可以确保掩蔽量化误差的频率子带中利用非零值进行编码。此外，可采用“频带分裂”方法，根据此方法，对于低端频率范围，前一层的输出信号完全不加修改，而增强层编解码器将只对该低端频率范围以上的一些附加高频分量编码。此方法尤其适用于包含语音的样本和适于执行语音编码的核

心层编解码器;

其次, 本发明可减少增强层编码过程中所需的比特数。甚至在最坏情况下, 与对一个含非零值的频率子带编码相比, 对一个只含零值的频率子带编码通常也会节省一些比特。在某些实验中发现, 根据本发明的这一方面, 可以节省大约总比特率的 10%。此第二优点可用于降低增强层编解码器的比特率或减少同一帧中其它频率子带的量化误差。

除了上述发明内容, 本发明人设想, 如果将编码器和解码器设计成以有效方式分别对零值进行编码和解码, 就可对音频信号进行更快的编码和解码。

参考如下详细公开、所附相关权利要求书以及附图, 可清楚本发明的其它目的、特征和优点。

附图说明

现在将更详细地描述本发明的优选实施例, 其中要参考附图, 附图中:

图 1 是可应用本发明的电信系统的示意图;

图 2 是说明图 1 所示系统的一些部件的示意框图;

图 3 是根据优选实施例的可伸缩音频编解码器的示意框图;

图 4 显示了一种示例性谱音频帧格式, 它具有多个频率子带, 并可用于本发明;

图 5 显示了一种频率选择切换 (FSS) 数组, 它用于指示图 4 所示音频帧的不同频率子带的编码音频信号的来源;

图 6 简要地说明根据优选实施例, 以最小量化误差和最佳比特率对音频信号进行量化的主要步骤;

图 7 说明用于执行图 6 所示量化过程的改进的 AAC 速率/失真控制环路。

实施例的详细说明

图 1 是可应用本发明的电信系统的示意图。在图 1 所示系统中，可以借助不同的网络 110、120 和 130 在不同的单元 100、112、122 和 132 之间传送音频数据。音频数据可以表示语音、音乐或任何其它类型的声学信息。因此，语音可以从固定电话 132 的用户通过公用交换电话网（PSTN）130 和移动通信网 110，经移动通信网 110 的基站 104，通过无线通信链路 102 传送到移动电话 100，反之亦然。该移动电话可以是市场上可购得的用于任何已知移动通信系统，如 GSM、UMTS 或 D-AMPS 的装置。

而且，存储在数据库 124 中的数字化的编码音乐可以从服务器 122 经因特网 120 和移动通信网 110 传送到移动电话 100 或可访问移动通信网 110 的另一便携式装置 112。便携式装置 112 可以是例如个人数字助理、具有 GSM 或 UMTS 接口的膝上型计算机、智能手持机或此类装置的另一附件等。服务器 122 所提供的音频数据可以不存储在数据库 124 中，而是直接从光学存储装置，如 CD 或 DVD 中读取。而且，服务器 122 可以连接到或包括于无线电广播台中，以便通过因特网 120 向便携式装置 100、112 提供流式音频服务。

因此，图 1 所示的系统仅用于举例，在本发明范围内，也可能存在在不同单元之间传送音频数据的其它各种情况。

图 2 是移动音频数据传输系统的一般框图，包括移动终端 250 和网站 200。例如，移动终端 250 可以表示图 1 所示的移动电话 100，而网站 200 可表示图 1 中移动通信网 110 的基站 104。

移动终端 250 可通过传输信道 206（例如图 1 中的移动电话 100 和基站 104 之间的无线链路 102）传送到网站 200。麦克风 252 接收来自移动终端 250 的用户的语音输入并将该输入转换为对应的模拟电信号，此信号提供给音频编码/解码单元 260。此单元含音频编码器 262 和音频解码器 264，它们共同构成音频编解码器。模拟麦克风信号经滤波、采样和数字化，然后由音频编码器 262 执行适用于

移动通信网的音频编码。音频编码/解码单元 260 的输出提供给信道编码/解码单元 270，其中，信道编码器 272 将根据移动通信网中适用的标准对编码音频信号执行信道编码。

5 信道编码/解码单元 270 的输出提供给射频 (RF) 单元 280，该单元包括 RF 发送器 282、RF 接收器 284 以及一个天线（未在图 2 中示出）。如本技术领域所熟知，RF 单元 280 包括各种电路，如功率放大器、滤波器、本地振荡器和混频器，它们将共同把编码音频信号调制到载波上，载波以电磁波形式从移动终端 250 的天线发射出去。

10 在经信道 206 传送之后，传输的 RF 信号及其中所含的编码音频数据由网络站 200 中的 RF 单元 230 接收。类似于移动终端 250 中的单元 280，RF 单元 230 包括 RF 发送器 232 及 RF 接收器 234。接收器 234 以基本上与上述发送器 282 所执行的过程相反的方式接收并解调所接收的 RF 信号，并将输出提供给信道编码/解码单元 220。信道解码器 224 将接收信号解码并把输出提供给编码/解码单元 220，
15 其中，音频解码器 214 对原来由移动终端 250 中的音频编码器 262 编码的音频数据进行解码。经解码的音频输出 204（例如 PCM 信号）可以在移动通信网 110 内转发（以传送到系统中所包括的另一移动终端）或者可以转发到例如 PSTN 130 或者因特网 120。

20 当音频数据沿相反方向传送，即从网络站 200 到移动终端 250 传送时，由音频编码/解码单元 210 的音频编码器 212 接收来自例如服务器 122 或固定电话 132 的音频输入信号 202（如 PCM 信号）。在该音频输入信号应用编码之后，由信道编码/解码单元 220 中的信道编码器 222 执行信道编码。然后，由 RF 单元 230 中的发送器 232
25 将编码音频信号调制到载波上，经信道 206 传送到移动终端 250 中的 RF 单元 280 的接收器 284。接收器 284 的输出提供给信道编码/解码单元 270 的信道解码器 274，在其中解码并转发给音频编码/解码单元 260 的音频解码器 264。音频数据经音频解码器 264 解码，最终

转换成模拟信号，该模拟信号经过滤波提供给麦克风 254，麦克风 254 将把传来的音频信号以可闻的形式呈给移动终端 250 的用户。

众所周知，移动终端 250 的音频编码/解码单元 260、信道编码/解码单元 270 以及 RF 单元 280 的操作由控制器 290 控制，控制器 290 具有相关联的存储器 292。对应地，网络站 200 的音频编码/解码单元 210、信道编码/解码单元 220 以及 RF 单元 230 的操作由具有相关存储器 242 的控制器 240 控制。

图 3 更详细地显示了图 2 所示的音频编码器 262。在优选实施例中，音频编码器 262 包括 AMR-WB 核心层编解码器 304（属于 CELP 类型）以及 AAC 增强层编解码器（由图 3 所示的大多数部件构成）。优选实施例的增强层是 MPEG-4 AAC 编解码器；但在本发明范围内，可将 MPEG-2 AAC 编解码器、ISO/MPEG 音频层-3（MP3）编解码器或任何其它采用例如离散余弦变换或小波变换工作的频率变换编解码器用作增强层编解码器。如何选择核心层编解码器不是本发明的中心；可以不采用 AMR-WB CELP 编解码器 304 而采用其它各种已知的编解码器（最好是但不一定是 CELP 类型），包括但不限于 MP3、AAC、AMR-NB（自适应多速率窄带）或 EFR（增强全速率）编解码器。

包括 CELP 核心层编码以及 AAC 增强层编码的可伸缩音频编码实际上是本技术领域所熟知的；因此这里无需作详细说明。可参考例如 ISO/IEC 14496-3:2001(E)子部分 4。此外，AMR-WB CELP 编解码器在第三代合作计划（3GPP）TS 26.190 V5.0.0（2001-03）中有描述。因此，为清楚起见，图 3 不是 AMR-WB CELP 核心层和 AAC 增强层音频编码器的完整示意图，而是用于说明本发明的核心方面。因而图 3 中已省略了 AAC 增强层编解码器的一些音频编码部件，例如：

TNS（时间噪声整形）。此部件对待编码音频信号的时间包络进行平滑处理，以控制编码噪声的精细时间结构。

LTP（长期预测）和预测。此部件通过预测减少音频信号的冗余度。

PNS（可察觉噪声替换）。此部件提供类噪声信道的有效表示。

5 M/S（中/侧立体声）。此部件增强立体声的镜像质量，并在一定程度上改善编码效率。

10 如图所示，音频编码器接收数字非压缩音频输入信号 300，如 PCM 信号或者本质上可以是本技术领域已知的其它任何数字音频信号。所述音频信号可以例如由图 1 中的音乐或无线电广播流式服务器 122 提供，由 PSTN 130 以语音信号的形式提供，如此等等。或者如本技术领域技术人员容易理解的那样，音频编码器可适于接收来自例如麦克风的模拟电信号，并通过常规的 A/D 转换将该模拟信号转换成数字信号。

15 在优选实施例中，逐帧执行音频编码，其中，每个帧在预定时间间隔，如 10 毫秒内包括多个音频样本。相邻帧的音频样本可以是不重叠的或者是部分重叠的。

20 输入信号 300 提供给感觉模型 330，其用途下面将会予以说明。此外，输入信号 300 还以如下方式提供给核心层编解码器 304。首先，在部件 302 中对输入信号 300 进行向下采样（down sample），以将采样频率变换为适合于核心层编解码器 304。随后，在部件 306 中执行核心层编码，并在部件 312 中将核心层输出 310 连同增强层编解码器的输出 370 多路复用，以形成编码音频输出流 314，随后在图 2 所示的部件 222 中对该输出流进行信道编码。

25 此外，核心层编解码器 304 在部件 308 中对编码的核心层信号解码并在部件 316 中对解码结果进行向上采样（up sample），以便得到重建的核心层信号 318，此信号是经核心层编码和解码之后的信号表示。类似于所有感觉音频编码，核心层编解码器 304 会在音频数据中引入一些失真。因此，重建的核心层信号 318 不会与原输入信号 300 相同。

原输入信号 300 在滤波器组 320 中变换成频域中的对应信号 324。相应地，重建的核心层信号 318 在滤波器组 322 中变换到频域。滤波器组 320 和 322 即便在图 3 中显示为分立的部件，但它们可以实现为一个公用的滤波器组，该滤波器组应用符合 MPEG-4 标准的改进的离散余弦变换 (MDCT)。

如图 3 中虚线 328 所示，必要时滤波器组 320、322 可以可选地由感觉模型 330 的输出控制，以便缩减滤波器组的窗口长度。滤波器组 320 和 322 中的 MDCT 变换可生成具有多个频率子带的谱音频帧。图 4 显示了这种音频帧的一个示例 400，它具有 49 个子带，即 401、402、...、449，共有 1024 个 MDCT 系数。这些子带可以是部分重叠的，或者可以是不重叠的。如图 4 所示，较低频率子带每个子带所具有的 MDCT 系数比较高频率子带少。经变换的原始信号 324 馈送给频率选择切换单元 (FSSU) 332 以及减法单元 334，减法单元 334 还接收经过变换的核心层信号 326。减法单元 334 通过从原始信号 324 的 MDCT 系数中减去核心层信号 326 的 MDCT 系数而得到剩余信号 336。剩余信号 336 提供给 FSSU 332，作为由核心层编解码器 304 引起的误差的指示。

对于谱音频帧 400 的每个子带 401-449，FSSU 332 分别计算原始信号 324 和剩余信号 332 的感觉熵 338、340。涉及对各子带编码所需的估计比特量的感觉熵可以计算为：

$$\text{感觉熵} = \frac{\text{信号能量}}{\text{掩蔽门限}}$$

其中，可以根据本技术领域熟知的各种方式中的任何一种来计算信号能量。掩蔽门限由感觉模型 330 提供，表示这样一种极限，在该极限以下，子带内容对人耳而言是不可闻的。掩蔽门限也可以各种方法来确定，其中一种在 Wang Y.和 Vilermo M.所著的“An Excitation Level Based Psychoacoustic Model for Audio Compression” (the 7th ACM International Multimedial Conference, October 30 to

November 4, 1999 Orlando, Florida, USA) 论文中描述, 该论文这里通过引用结合于本文中。

5 在部件 342 中, FSSU 332 将计算得到的感觉熵 338、340 进行比较, 以确定信号 324 和 336 中哪一个信号具有最小的感觉熵 (并因而在具有同样质量的前提下需要较少的比特来编码), 并相应地将图 5 所示的 FSS 数组 500 中的各控制信息标志 501-549 置位。因此, 针对不同的子带 401-449, 根据 FSSU 332, 在部件 344 和 346 中分别选用原始信号 324 或剩余信号 336, 从而分别形成相应的输出 360 和 362。音频帧 400 的每个子带 401-449 具有各自相应的标志 10 501-549, 各标志最好用一个二进制位表示。FSS 数组将与编码音频数据 310、370 一起在多路复用比特流 314 中传送到接收端, 以在解码器对收到的编码音频数据解码时向其指示是否要将增强层信号加到核心层信号中。

15 通过将各频率子带 401 - 449 的原始信号 324/360 或剩余信号 336/362 的 MDCT 系数拷贝到一个具有图 4 所示格式的公共帧数组中, 根据 FSSU 数组 500 中的比特位设置, 从而组成 FSSU 332 的输出。所组成的输出对应于普通的 AAC 增强层编解码器, 被提供给速率/失真控制过程 348, 此过程将在后面参照图 6 和图 7 作更详细的说明。

20 除了以上所述, 根据优选实施例, 还针对当前的频率子带 401-449, 并行地将剩余信号 336 与掩蔽门限 350 作比较, 如部件 356 所示。如果确实发现剩余信号低于掩蔽门限, 则这意味着核心层编解码器 304 引起的误差在当前频率子带内是不可闻的。因此, 在这种情况下, 在此特定频率子带中只将零值作为增强信号编码就已经足 25 够了, 如部件 358 和 364 所示。此条件具有最高优先级, 可使感觉熵计算的输出 360/362 的控制作用无效, 从而还有利于稍微加快后续量化过程的计算。

此外, 在部件 352 和 354 中, 针对每个频率子带 401-449 计算

掩蔽门限 350 和剩余信号 334 之间的差值。为了提高精度，可以将快速傅立叶变换（FFT）而非 MDCT 系数用于此计算。其结果对每帧 400 累加，以反映核心层质量的量度，即指示核心层编解码器 304 在核心层对音频输入信号 300 编码时性能的好或坏。如后面将要描述的那样，对于每个帧，核心层质量量度将用作乘法因子（参见图 7 中的步骤 710），在部件 348 所示的后继量化过程中先于量化误差与剩余信号作比较之前应用于量化误差。

下面将参照图 6 和图 7，描述按照优选实施例作了改进的速率/失真控制过程 348。该过程基于 AAC 中的普通速率/失真控制过程，AAC 涉及比例因子单元、量化单元和无噪声编码单元，这些单元在所述 ISO/IEC 14496-3: 2001 (E) 的附录 4.B 10-11 中有详细描述。

速率/失真控制过程的主要目的是为待编码音频信号提供量化和无噪声编码。根据优选实施例，量化过程 600 划分为一个两级过程：最初，对所有子带执行第一 AAC 量化过程 610。此第一 AAC 量化过程相对于标准的 AAC 量化过程作了改进，即考虑剩余信号而重复对量化误差进行评估，并可根据此评估结果决定对某些子带用零值来替换剩余信号。实际的替换仅在已处理所有子带之后才进行，如图 6 所示的步骤 620 所示。最后，对所有剩余子带（即非零值）执行第二 AAC 量化过程 630；不过这次无任何零值替换。

图 7 更详细地显示了 AAC 量化 610 的改进的速率/失真控制过程。由 FSSU 332 提供的信号 702 表示当前待量化的频域样本。在步骤 706 中，以不同的比例因子按不同方式将样本量化若干次。在每次量化之后，在步骤 708 计算量化误差并在步骤 718 中进行评估。每次，该过程都尝试改善量化噪声的“着色 (coloring)”，即该过程尝试以使量化噪声的可听度最小的方式将可用比特分配给不同的频带。这是通过针对不同频带改变比例因子来完成的。比例因子（一个频带对应一个）控制量化步长大小，从而将不同的噪声量分配给各频带。在每轮量化之后，检查每个频带中的噪声，对于其中噪声

最令人讨厌的频带，将在下一轮中减小其相应量化步长。

但是，此过程不一定会收敛到一个全局最优的情况，因此，对于某些频带，核心编解码器可造成比量化的剩余信号（quantized residual）小的误差。对于这些频带，随后仅使用核心编解码器是有利的。要注意的是，要量化的信号可以是原始信号或剩余信号（从原始信号中减去核心编解码器信号所得的信号）。

因此，以下参照图 7 中的步骤 710-720 来执行。在步骤 714 中要检查量化噪声和仅使用核心编解码器会得到的噪声（即 FSSU 332 提供的剩余信号 704），而不是在每轮量化之后只检查量化噪声。如果核心编解码器输出足够好，则无需减小此频带的量化步长大小。如果量化噪声和核心编解码器噪声均太高，则减小步长大小。

如果核心编解码器总的来说表现良好（与语音信号和语音核心编解码器的情况一样），则最好是更进一步地修改误差计算。如果核心编解码器表现良好，这将反映在核心层质量量度 710 的取值上，该核心层质量量度 710 是在图 3 中的部件 354 中导出的。在这种情况下，可以忽略核心编解码器信号中的小误差，并且仍然可以仅使用核心编解码器信号，即便只使用核心编解码器引起的误差将比量化误差稍微大一些。使用其中核心编解码器误差稍微大于量化误差的频带中的核心编解码器信号将获得可供其中核心编解码器误差比量化误差大得多的其它频带使用的比特。因此，在图 7 所示步骤 712 中将步骤 708 中形成的量化误差与核心层质量量度的函数值相乘，然后在步骤 714 中将相乘结果与核心层噪声作比较。该函数可以例如基于简单门限。如果核心编解码器表现良好，则核心层质量量度保持在门限（例如为 0）之上，且量化误差与预定常数如 4 相乘。否则量化误差在步骤 714 中的比较中保持不变。

这样形成的量化循环具有三个中断条件：

1. 无可察觉量化误差出现；
2. 进一步减少量化误差是不可能的；以及

3. 已达到最大循环次数极限。

在运行几轮量化之后，将满足所述中断条件。如果在某些频带中使用核心编解码器输出而非量化结果仍然较好，则在这些频带中以零值来替换量化信号，如步骤 720 所示。此外，将标记对应这些频带的 FSS 信息（参见步骤 716），以指示要使用剩余信号，从而有效地使解码器在这些频带中只使用核心编解码器信号（在最后一种情况中，将不重新分配比特，但所述方法可节省比特）。

如果存在用零值替换过的频带，则量化信号需要少于所计算的比特的比特。通过在各帧之间使用比特缓冲器，这些比特可留给后续帧使用。这些比特还可用于改善当前帧的量化，方法是如图 6 中的步骤 630 所示那样再次运行量化循环，但这次从第一次（即在图 6 中步骤 610 中）取得的最终结果开始，并如上所述包括某些频带中的零值替换。在第二轮量化过程 630 中，对信号进行量化，同时无需进行就第一轮量化过程 610 所述的误差计算修改，即步骤 630 中的量化完全符合 MPEG-4 AAC 标准。

作为上述的一种替代选择，可以在一个公共量化过程中决定用零值替换哪些子带，以及确定剩余（非零）子带的比特率，而不是对所有子带执行第一量化过程，用零值替换相关子带且随后对剩余子带执行附加的量化过程。

根据本发明的音频编码器可有利地包括在例如 GSM 或 UMTS 网络中的音频码变换器中。在 GSM 中，这种音频码变换器称为码变换/速率适配单元（TRAU），并在来自 PSTN 130 的 64 kbps PCM 语音与全速率（FR）或增强全速率（EFR）13-16 kbps 数字化 GSM 语音之间提供转换。音频码变换器可位于基站收发信台（BTS）处（为基站子系统（BSS）的组成部分），或者位于移动交换中心（MSC）处。

上述的可伸缩音频编码功能可以实现为集成电路(ASIC)或任何其它形式的数字电路。在替代实施例中，所述可伸缩音频编码功能

可以实现为计算机程序产品，该程序产品可以直接装入处理器的存储器中-优选图 2 所示的网络站 200/移动台 250 的控制器 240/290 及其相关存储器 242/292。该计算机程序产品包含在由所述处理器执行时用于提供所述可伸缩编码功能的程序代码。

5 根据本发明的音频编码器还可以包括在用于向网络客户（如图 1 中的便携式装置 100、112）提供音频服务的网络服务器（如图 1 中的服务器 122）中。

如前所述，不必对以上本发明实施例中的解码器侧作任何改变，因为该解码器将对标记在 FSS 数组 500 中的每个子带 401-499 中的
10 在先层输出信号的 MDCT 系数求和。但是，本发明的替代实施例包括一个改进的解码器，该解码器适于接收包含编码的第一层信号和至少一个已编码的第二层信号的分层数据流。该数据流将包含编码帧，该编码帧具有多个谱子带，但另外可能具有任意已知的格式。在已将第一层信号和第二层信号二者解码之后，解码器分析第二层
15 信号并确定每个子带的感觉质量量度。随后，解码器判断该感觉质量量度是否满足预定的判据，如果满足，则第二层信号与所述子带的第一层信号合并，以形成解码输出信号。否则，只根据该子带的第一层信号生成解码输出信号，即以零值替换第二层信号。

感觉质量量度可以包含在收到的分层数据流，即源自编码器侧
20 的分层数据流中。或者，解码器可以通过在已解码的第二层信号中识别非自然音频分量，如毛刺、噪声或非预期的波形而导出感觉质量量度，并因此决定已解码的第二层信号是失真的，不应加到解码的第一层信号中。

以上主要参照一个实施例对本发明作了描述。但是，如本领域的
25 技术人员容易理解的那样，在本发明范围内存在与以上所述实施例不同的同样可行的其它实施例，这些实施例由所附权利要求书限定。

要强调的是，本发明不限于一个基础层和一个增强层；本发明

原理同样可适用于多层编码技术中的两个相继增强层。此外，在替代实施例中，可只使用一些而非全部频带的 FSS 信息并将其发送到接收侧。这在仅将增强层用于改善高频的情形中可能有用。因此，无需发送低频的 FSS 信息。当然，必须以缺省方式或通过发送侧的初始握手方式将此情况通知接收侧。

此外，以上所述的音频编码器、解码器、码变换器、计算机程序和集成电路不限于以上所述的位置。例如，音频编码器可以位于图 1 所示的音频发送装置中的任意位置上，或者位于未示出或未描述的另一音频发送装置中。

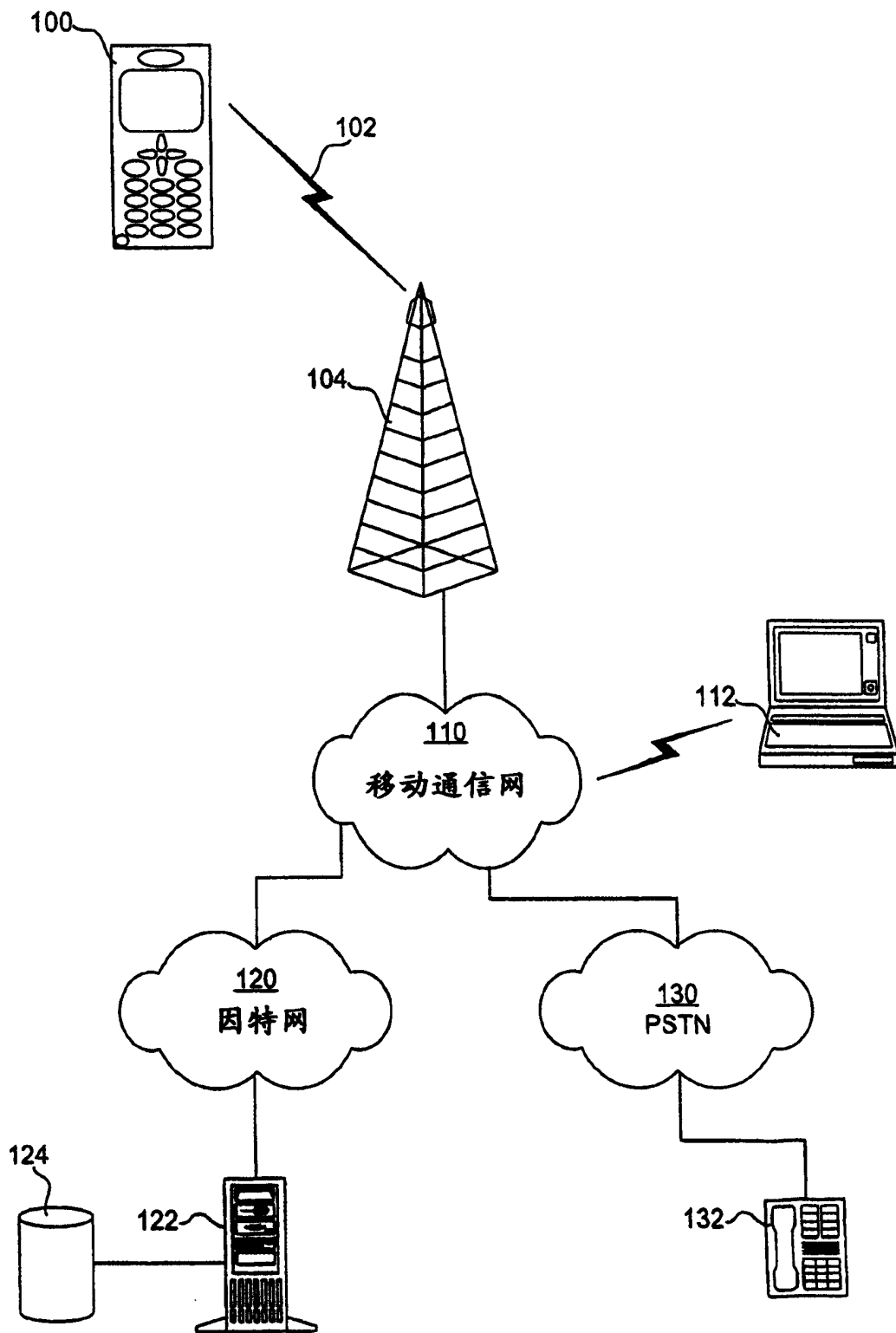


图 1

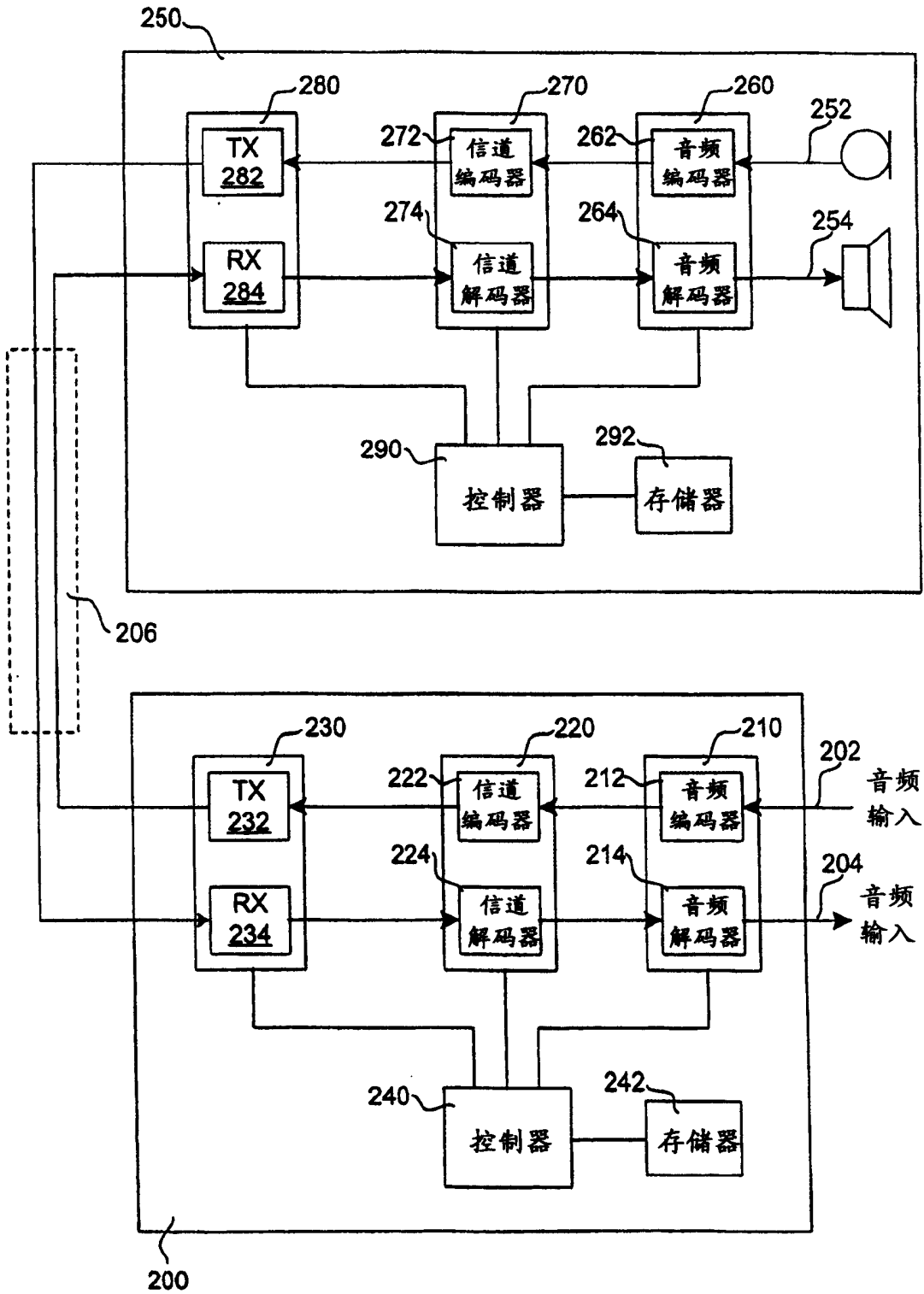


图 2

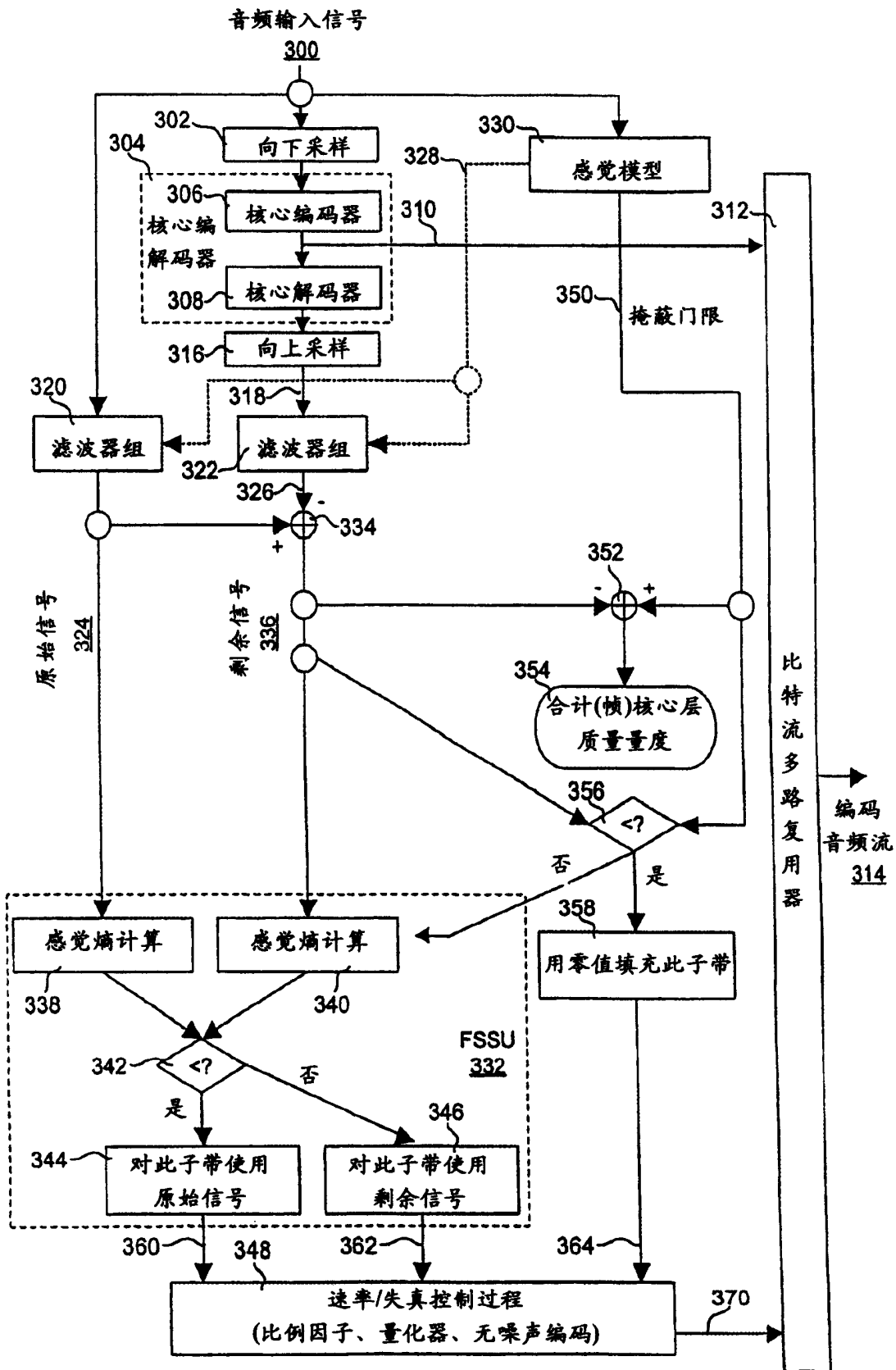


图 3

图 4

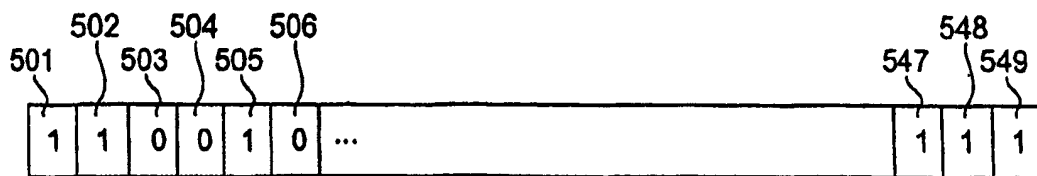
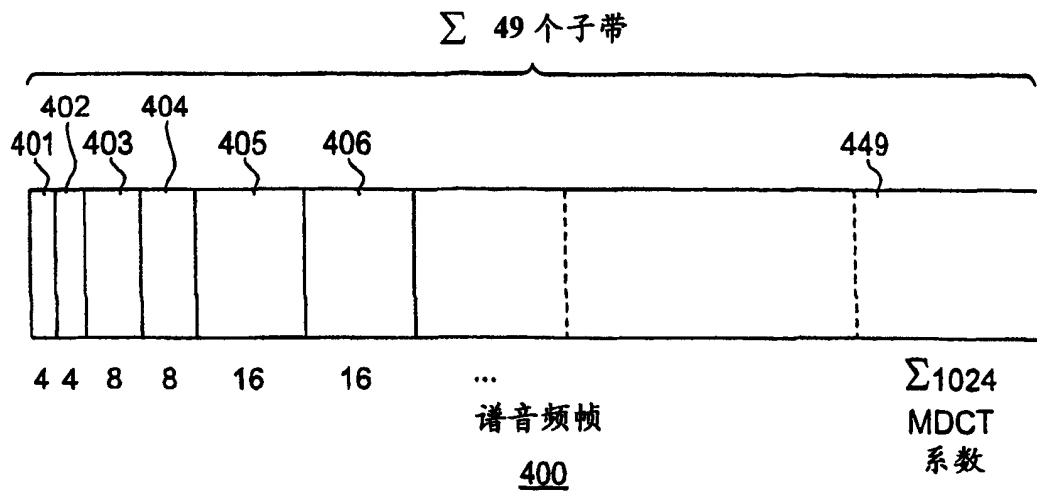


图 5

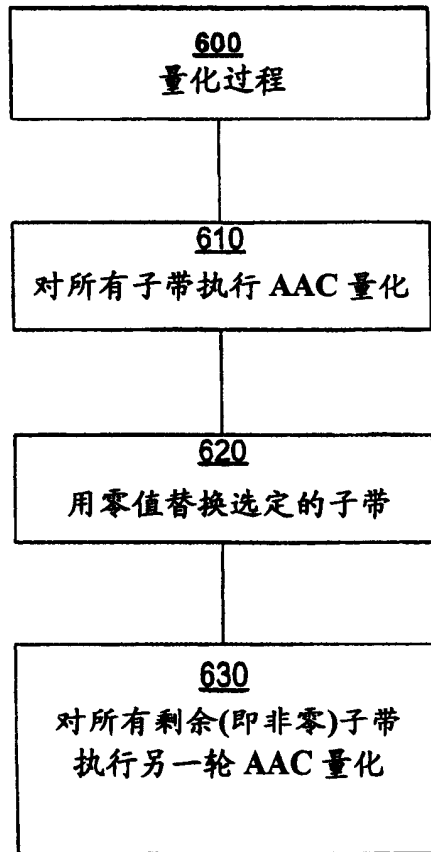
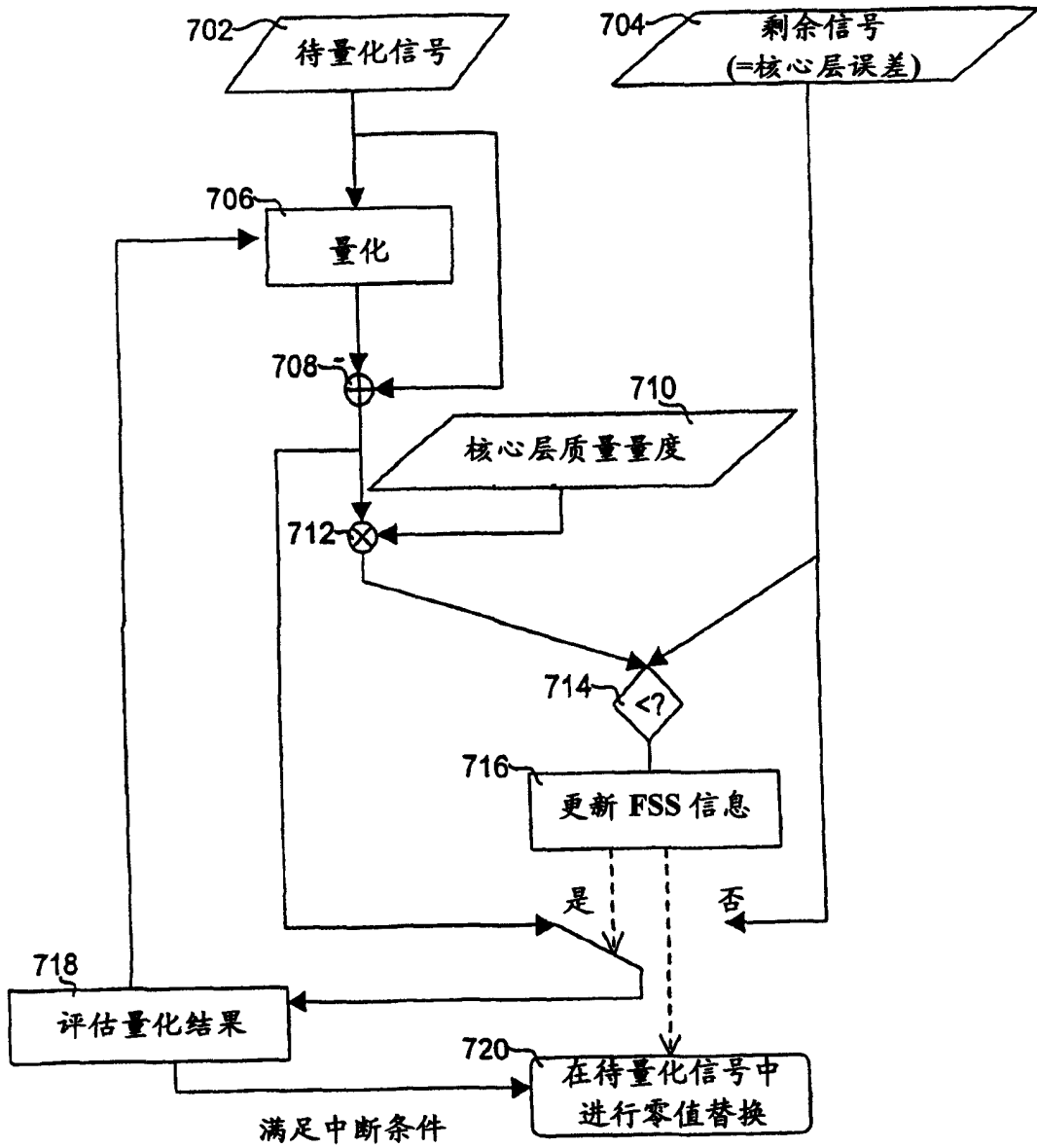


图 6



改进的速率/失真控制环路

图 7