



(12) 发明专利

(10) 授权公告号 CN 101872618 B

(45) 授权公告日 2012.08.22

(21) 申请号 201010126591.9

(51) Int. Cl.

(22) 申请日 1996.11.21

G10L 19/02 (2006.01)

(30) 优先权数据

60/007,896 1995.12.01 US

08/642,254 1996.05.02 US

(56) 对比文件

WO 9502925 A1, 1995.01.26, 全文.

WO 9410772 A1, 1994.05.11, 全文.

(62) 分案原申请数据

96199832.6 1996.11.21

审查员 张鑫

(73) 专利权人 DTS(BVI) 有限公司

地址 英属维尔京群岛托尔托拉岛

(72) 发明人 斯蒂芬·M·史密斯

迈克尔·H·史密斯

威廉·保罗·史密斯

(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 11038

代理人 李镇江

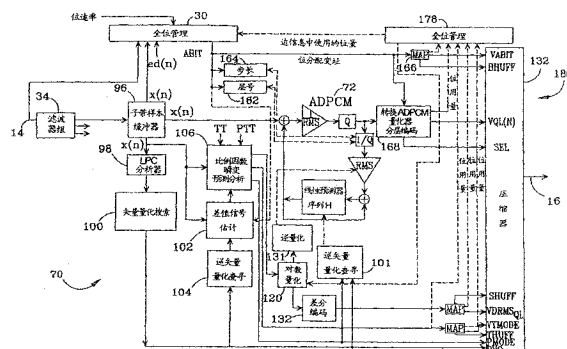
权利要求书 3 页 说明书 26 页 附图 18 页

(54) 发明名称

多通道音频编码器

(57) 摘要

本发明涉及一种多通道音频编码器。一种子带声编码器 (12) 采用了完全 / 非完全重构滤波器 (34)、预测 / 非预测子带编码 (72)、瞬变分析器 (106)、和心理声学 / 最小均方差 (mmse) 相对时间的位分配 (30)、频率以及多声道对数据流进行编码 / 解码以产生高保真度的重构声音。声编码器 (64) 对多通道声信号进行划分以便把帧尺寸即字节数量限制在所需的范围内, 并对编码数据进行格式化处理以便当接收到各子帧时能进行播放从而减少伪象。此外, 声编码器处理声频带宽的基带部分 0-24kHz 以使用相同的编码 / 解码算法对 48kHz 或更高的频率进行采样从而使声编码器的结构在未来具有兼容性。



1. 一种多通道音频编码器,以已知比特率编码多通道音频信号,包括:

取帧器,其向以某一采样率采样的多通道音频信号的每个通道应用音频窗口以产生相应的音频帧序列;

多个滤波器,均包括非完全和完全重构滤波器,在已知比特率分别低于和高于阈值比特率时,在基带频率范围内把音频帧分成相应的多个频率子带,所述频率子带中的每一个包含有子带帧序列,在每个子带帧中具有至少一个音频数据子帧;

多个子带编码器,一次一个子帧地将相应频带中的音频数据编码为经编码的子带信号;以及

多路复用器,其针对每个连续数据帧将经编码的子带信号和滤波器选择码打包和多路复用成输出帧,由此形成具有某一传输率的数据流。

2. 一种对以某一采样率采样的多通道音频信号进行编码的方法,包括:

应用音频窗口到多通道音频信号的每个通道以产生相应的音频帧序列;

在基带频率范围内将通道的音频帧分成相应的多个频率子带,所述频率子带中的每一个包含子带帧的序列,每个子带帧具有至少一个音频数据子帧,其中每个子帧包括至少一个子子帧;

一次一个子帧地将相应频率子带中的音频数据编码成经编码的子带信号;以及

针对每个连续音频帧将经编码的子带信号多路复用成输出帧以产生具有某个传输率的数据流,其中音频窗口的大小根据传输率与采样率的比值被选择以便使得所述输出帧的大小被限制在期望的范围内,所述输出帧的大小、子帧的数量以及子子帧的数量被多路复用为所述输出帧。

3. 如权利要求2所述的方法,其中经编码的子带信号一次一个子帧地被打包为输出帧,其中它们自己的辅助信息包括位分配值,使得每个连续子帧不用参考任何其它子帧就可解码。

4. 如权利要求2所述的方法,其中多路复用步骤将子帧结束码插在每个子帧的末尾以提供差错校验。

5. 如权利要求2的方法,其中编码频率子带的步骤包括:

将每个子帧分成多个子子帧;

为该子帧产生估计的差信号;

在估计的差信号的每个子子帧中检测瞬变;

产生瞬变代码以指出是否在除第一子子帧之外的任何其它子子帧内存在有瞬变以及瞬变出现在哪个子子帧内;

在检测到瞬变时,为该瞬变之前的那些子子帧产生一个预瞬变比例因数,并且为包括该瞬变和该瞬变之后的那些子子帧产生一个后瞬变比例因数,否则,为该子帧产生一个统一比例因数;

为当前子帧产生差信号;

根据预瞬变比例因数、后瞬变比例因数以及统一比例因数对差信号进行比例运算;以及

在当前子帧上以固定比特率量化经比例运算的差信号。

6. 一种对以某一采样率采样的多通道音频信号进行编码的方法,包括:

向多通道音频信号的每个通道应用音频窗口以产生相应的音频帧序列；

在基带频率范围内将通道的数据帧分成相应的多个频率子带，所述频率子带中的每一个包含子带帧的序列，每个子带帧具有至少一个音频数据子帧；

通过以下步骤为音频窗口中的子帧产生比特分配值：

针对每个子帧从所述音频数据中产生估计的差信号；

基于所述估计的差信号为每个子帧计算心理声学信号 - 掩蔽比值 SMR；

分配比特以满足每个子帧的信号 - 掩蔽比值；

为所有子帧计算分配的比特率；以及

当分配的比特率低于目标比特率时，根据最小均方差 mmse 方案为子帧分配剩余比特；

利用预测编码根据所述比特分配值一次一个子帧地编码从相应频率子带中的音频数据导出的差信号以产生经编码的子带信号；以及

对于每个连续数据帧多路复用经编码的子带信号为输出帧，以产生具有某个传输率的数据流。

7. 如权利要求 6 所述的方法，其中利用预测编码器编码频率子带，所述产生比特分配值的步骤还包括：

为每个子帧计算估计的预测增益；以及

将信号 - 掩蔽比值减小其相关的估计的预测增益的相应部分。

8. 如权利要求 6 所述的方法，其中分配剩余比特的步骤包括：

为每个子帧计算均方根 RMS 值；

根据应用于均方根值的最小均方差方案分配所有的可用比特，直到分配的比特率接近目标比特率。

9. 如权利要求 6 所述的方法，其中分配剩余比特的步骤包括：

为每个子帧计算均方根 RMS 值；

根据应用于均方根值的最小均方差方案分配所有的剩余比特，直到分配的比特率接近目标比特率。

10. 如权利要求 6 所述的方法，其中分配剩余比特的步骤包括：

为每个子帧计算均方根 RMS 值；

根据应用于子帧的均方根值和信号 - 掩蔽比值之差的最小均方差方案分配所有的剩余比特，直到分配的比特率接近目标比特率。

11. 一种多通道音频编码器，包括：

取帧器，其向以某一采样率采样的多通道音频信号的每个通道应用音频窗口以产生相应的音频帧序列；

多个滤波器，其在基带频率范围内将各通道的数据帧分成相应的多个频率子带，所述频率子带的每一个包含子带帧序列，每个子带帧具有至少一个音频数据子帧；

分析器，其产生估计的误差信号，从该误差信号为每个子帧计算预测增益；

多个子带编码器，其将相应通道的频率子带中的音频数据编码为经编码的子带信号，每个子带编码器包括编码较低频率子带的多个自适应差分脉冲编码调制 ADPCM 编码器，所述 ADPCM 编码器的预测能力在它们相应的预测增益低于阈值增益时被禁用，从而形成编码较低频率子带的自适应脉冲编码调制 APCM 编码器；和编码较高频带的多个矢量量化器 VQ；

以及

多路复用器,其针对每个连续数据帧将经编码的子带信号打包和多路复用为输出帧,从而形成具有某个传输率的数据流。

12. 如权利要求 11 所述的多通道音频编码器,其中矢量量化器编码高于阈值频率的所有子带。

13. 如权利要求 11 所述的多通道音频编码器,其中所述基带频率范围具有最大频率,所述多通道音频编码器还包括:

预滤波器,其将每个所述音频帧分别分成在基带频率范围内的频率处的基带信号和在最高频率之上的高采样率信号;以及

高采样率编码器,其将音频通道的高采样率信号编码为相应的经编码的高采样率信号,

所述多路复用器将各通道的经编码的高采样率信号打包成相应的输出帧,使得多通道音频信号的基带和高采样率部分可独立解码。

14. 如权利要求 11 所述的多通道音频编码器,其中多通道音频信号以目标比特率编码,该多通道音频编码器还包括:全局比特管理器 GBM,其为每个子帧计算心理声学信号-掩蔽比值 SMR,并且在启用 ADPCM 编码器时,利用信号-掩蔽比值相关的预测增益的相应部分修正信号-掩蔽比值,分配比特以满足每个信号-掩蔽比值,计算整个子带内的分配的比特率,以及调整各个分配值,使得实际比特率接近目标比特率。

15. 如权利要求 14 所述的多通道音频编码器,其中子带编码器将每个子帧分成多个子子帧,该多通道音频编码器还包括:

分析器,其在启用 ADPCM 编码器时在每个子子帧中检测估计的误差信号中的瞬变,以及在启用 APCM 编码器时检测音频数据中的瞬变;产生瞬变代码以指出是否在除第一子子帧之外的任何其它子子帧内存在有瞬变以及瞬变出现在哪个子子帧内,以及在检测到瞬变时,为该瞬变之前的那些子子帧产生一个预瞬变比例因数,并且为包括该瞬变和该瞬变之后的那些子子帧产生一个后瞬变比例因数,否则,为该子帧产生一个统一比例因数,

在进行编码以降低对应于预瞬变比例因数的子子帧中的编码误差之前,所述 ADPCM 和 APCM 编码器分别利用所述预瞬变、后瞬变和统一比例因数对误差信号和音频数据进行比例运算。

多通道音频编码器

[0001] 本分案申请是基于申请号为 96199832.6, 申请日为 1996 年 11 月 21 日, 发明名称为“多通道声解码器”的中国专利申请的分案申请。更具体说, 本分案申请是基于申请号为 200610081786.X, 申请日为 1996 年 11 月 21 日, 发明名称为“多通道音频编码器”的分案申请的再次分案申请。

技术领域

[0002] 本发明涉及多通道音频信号的高质量编码和解码, 更确切地说, 是关于一种子带编码器, 此编码器在时间域、频率域和多个音频通道之间运用了完全 / 非完全重构滤波器组、预测 / 非预测子带编码、瞬态分析、和心理声学 / 最小均方差 (MMSE) 比特率分配等方法, 以产生可使其相应解码计算量受到约束的数据流。

背景技术

[0003] 公知的高质量音频和音乐编码器可以分成两大类方案。第一类是具有中高频率分辨率的子带 / 变换编码器, 该类编码器能根据心理声学的掩蔽计算结果自适应地量化其分析窗内的子带或系数采样数据。第二类是频率分辨率较低子带编码器, 这种编码器通过 ADPCM (自适应差分脉冲编码调制) 对子带采样数据进行处理来补偿其频率分辨率的不足。

[0004] 第一类编码器利用了音乐信号中大量的短时频谱变化差异, 使其比特率分配根据信号的频谱能量进行自我调整。由于其频率分辨率高的特性, 这些编码器变换后的频域信号能被直接应用于建立在听觉的临界频带理论上的心理声学模型。托德 (Todd) 等人在 1994 年 2 月发表于音响工程协会年会的“AC-3: 音频传输和存储的灵活感知型编码”一文中的 Dolby AC-3 音频编码器, 就典型地对各个 PCM 信号进行 1024-点 ffts (快速傅立叶变换) 计算, 并将心理声学模型应用于每个通道的 1024 个频率系数以决定其比特率。Dolby 系统并将窗口尺寸减小到 256 个采样以隔离信号的瞬态响应, 进行瞬态分析。AC-3 编码器采用了专用的后向自适应算法对比特率分配信息进行解码。这样便减少了随编码音频数据一同发送的比特率分配信息量。其结果是, 相对于前向自适应方法来说, 可用于音频的带宽得以增加, 从而改善了音质。

[0005] 在第二类编码器中, 子带差分信号或是被固定量化, 或是在量化时可动态调整以使所有或部分分段频带上的量化噪声减至最小, 它们并没有明确参照心理声学掩蔽理论。由于在码率分配处理之前很难估计预测器性能, 通常认为不能将心理声学失真阈值直接应用到预测 / 差分子带信号上。而量化噪声对预测过程的反作用使问题进一步复杂化。

[0006] 这类编码器能有效工作是因为听觉感知中重要的音频信号通常在长时间段中展示出的周期性特征。这种周期性可被预测差分量化过程充分利用。把信号分成少数子带可降低可以听觉的噪声调制效应, 并可有效利用音频信号含有的长时频谱分量差异。然而随着子带数量的增加, 每个子带内的预测增益将不断降低, 增加到某一程度时预测增益将趋于零。

[0007] 数字影院系统公司 (DTS), L. P. 采用了一种音频编码器, 它将每个 PCM 声道过滤分

成四个子带,而且用后向 ADPCM 编码器对每个子带进行编码,且所述后向 ADPCM 编码器中的预测器系数可根据子带数据作自适应调整。编码器在每个声道上采用相同的固定比特码率分配,且低频子带较高频子带优先得以分配更多比特码率。固定码率分配方法提供了例如 4 : 1 的固定压缩比。迈克·史密斯 (Mike Smyth) 和斯蒂芬·史密斯 (Stephen Smyth) 在“APT-X100 :用于广播的低延时、低码率、子带 ADPCM 音频编码器”,第十次国际 AES 会议汇编,1991 年,第 41-56 页中描述了此类 DTS 编码器。

[0008] 此两类音频编码器还具有其它共同的局限性。首先,公知的音频编码器编码 / 解码时使用固定框架 / 帧尺寸,即采样数据数量或帧所占时间段是固定的。结果是,当编码的传输率相对于采样频率增加时,帧内的数据量也增加。因此,解码器缓冲器的尺寸必须设计得能容纳最坏情况以避免数据溢出。这样将会增加作为解码器主要成本成分的 RAM 的用量。其次,公知的音频编码器不易扩展应用于大于 48kHz 的采样频率。若这样做将出现已有解码器与新编码器所需格式不兼容的情形。缺乏未来兼容性是一个严重的局限性。此外,对 PCM 数据进行编码所用的公知格式要求解码器必须读入整个帧的数据之后才能开始播放。这又需要把缓冲器的尺寸限制在 100ms 左右的数据块,从而不产生过长的延时或滞后而干扰听者。

[0009] 此外,虽然这些编码器的编码能力高达 24kHz,但较高频子带却常常被割舍。这样会降低高频保真度或解译再造信号的听觉氛围。公知编码器通常使用两种错码检测方案中的一种。最常用的是里德所罗门编码 (Reed Solomon coding),其编码器将产生的检测码加入到数据流的辅助信息之中。这样便易于检测和纠正辅助信息中出现的任何错误。然而,它没有检测出音频数据中的错误。另一种方法是检验数据帧及其头信息区域是否无效码状态。例如,假设某个 3 比特参数只允许有 3 种有效状态。那么发现其它五种状态中的任何一种代表着错误的发生。这种方法只是提供了某种检测能力,而音频数据中的错误仍不能被发现。

发明内容

[0010] 鉴于上述问题,本发明提供了一种多通道音频编码器,其灵活性可容纳广大范围的压缩比要求,采用高比特码率时能产生比 CD 更好的质量,使用低比特码率时亦能改善听觉感知质量。它同时还具备降低播放延时、简化误差检测、改善预回音失真、及可延伸至未来更高采样率的特性。

[0011] 这是用子带编码器实现的,子带编码器把每个声道的音频信号用窗口法分成音频帧序列数据,然后将各帧数据进行滤波分成基带和高频区,再把每个基带信号分解成多个子带。子带编码器在码率较低时通常选择非完全滤波器以分解基带信号,而在码率足够高时选择完全滤波器。高频区信号独立于基带信号在高频编码阶段进行编码。基带编码阶段则包括分别用于较高频和较低频子带编码的 VQ 和 ADPCM 编码器。每个子带帧包括至少一个子帧,每个子帧进一步细分为多个子子帧。每个子帧被用作分析单元,以便估计 ADPCM 编码器的预测增益,且当预测增益较低时可终止使用其预测能力。子帧分析单元还被用来检测瞬变状态以调节瞬变状态前后的 SFs (比例因数)。

[0012] 全局比特管理 (GBM) 系统利用多个声道、多个子带和当前帧内各子帧之间的差别把比特码率按需分配到每个子帧。GBM 系统首先计算经过预测增益加以修正的 SMR (信号掩

蔽值比),并以心理声学模型为基础将比特码率分配到每个子帧。然后,GBM 系统按照 MMSE 方法分配所有剩余比特码率,它或者马上切换到 MMSE 分配法以降低总噪声底值,或逐渐趋变到 MMSE 分配分案。

[0013] 多路复用器产生输出帧数据,此输出帧包含有同步字节、帧头信息、音频头信息和至少一个子帧,并以传输速率用多路复用形式合并成为数据流。帧头信息包括窗口尺寸和当前输出帧的尺寸。音频头信息表示音频帧数据的打包安排和编码格式。每个音频子帧包括独立于其它子帧的音频解码辅助信息、高频 VQ 编码数据、多个基带音频子子帧(每个子子帧以多路复用形式包装着来自各声道的、较低频子带的音频数据)、一个高频音频数据块(以多路复用形式包装着来自各声道的、高频区域音频数据,用来支持多通道音频信号在解码时可取用多种高采样率)、以及用于核实检验子帧结束的拆包同步字节。

[0014] 窗口尺寸的选择决定于传输速率与编码器采样频率之比,从而把输出帧的尺寸限制在要求的范围内。当压缩量相对较低时,窗口尺寸减小从而使帧尺寸不会超过上限最大值。因此解码器可以采用相对较少的、固定数量的 RAM 作为输入缓冲器。当压缩量相对较高时,窗口尺寸增加。GBM 系统因此能利用较大的时间窗进行比特率分配,由此改善编码性能。

[0015] 对于熟悉本领域的技术人员来说,本发明的这些及其它特征和优点将会通过以下对优选实施例所作的详细说明而变得清晰。这些详细说明阐述将会结合附随图表,其中:

附图说明

- [0016] 图 1 是按照本发明所述的 5 声道音频编 / 解码器的框图 ;
- [0017] 图 2 是多通道编码器的框图 ;
- [0018] 图 3 是基带编码器和解码器的框图 ;
- [0019] 图 4a 和图 4b 分别是高采样率编码器和解码器的框图 ;
- [0020] 图 5 是单声道编码器的框图 ;
- [0021] 图 6 是使用不同传输率每帧字节与帧尺寸之间的关系图 ;
- [0022] 图 7 是 NPR(非完全)和 PR(完全)重构滤波器的幅值响应曲线图 ;
- [0023] 图 8 是重构滤波器的子带混叠示意图 ;
- [0024] 图 9 是 NPR 和 PR 滤波器的失真曲线图 ;
- [0025] 图 10 是单个子带编码器的示意图 ;
- [0026] 图 11A 和 11B 分别图解说明子帧中的瞬态检测和比例因数计算 ;
- [0027] 图 12 描述对量化后 TMODES 的熵编码过程 ;
- [0028] 图 13 描述比例因数的量化过程 ;
- [0029] 图 14 描述信号掩蔽曲线与信号的频率响应的卷积以产生 SMR ;
- [0030] 图 15 是人的听觉响应的曲线图 ;
- [0031] 图 16 是子带的 SMRs 曲线图 ;
- [0032] 图 17 是用于心理声学 and mmse 比特率分配的误差信号曲线图 ;
- [0033] 图 18A 和图 18B 分别是子带能量曲线图和其倒置曲线图,描述了 mmse“注水式”比特率分配过程 ;
- [0034] 图 19 是数据流中单帧结构的框图 ;

- [0035] 图 20 是相应解码器的示意图；
- [0036] 图 21 是编码器一种硬件实现方法的框图；和
- [0037] 图 22 是解码器一种硬件实现方法的框图。
- [0038] 附表说明
- [0039] 表 1 列出了使用各种采样频率和传输率时可取的最大帧尺寸；
- [0040] 表 2 列出了使用各种采样频率和传输率时可取的最大帧尺寸（字节单位）；
- [0041] 表 3 显示了 ABIT 索引值、量化级数量及所产生的子带 SNR（信噪比）之间的关系。

具体实施方式

[0042] 多通道音频编码系统

[0043] 如图 1 所示，本发明在其一体多通道音频编码器 10 中结合了两类已知编码方案的特征并加入了新的优点特征。所用编码算法是按照节目原制作室质量水平即“好于 CD”级质量而设计，其应用范围广泛，可满足在压缩量、采样频率、采样字长、声道数量及感知听觉质量等方面的不同要求。

[0044] 编码器 12 把通常在 48kHz 下以 16-24 位字长采样的多通道 PCM 声数据 14 编码成已知传输率的数据流 16，合适的传输率范围为 32-4096kbps。与已知的音频编码器不同是，本结构能扩展到更高的采样频率（48-192kHz），而不会使已有的、为基带采样频率或任何中间采样频率而设计的解码器出现不兼容的情况。此外，PCM 数据 14 被窗口法分成帧块逐一编码，而每个帧的优先选择是分成 1-4 个子帧。音频信号窗的尺寸，即其 PCM 采样数，则决定于相对的采样频率和传输率取值，它的选择使得输出帧的尺寸，即其相应解码器 18 读出的每帧数据字节数，被适当地限制在 5.3-8 千字节之间。

[0045] 其结果是，解码器中用于缓冲输入数据流的 RAM 数量可保持在较低水平，从而降低了解码器成本。低码率时可以使用较大的窗口尺寸将 PCM 数据构成帧，这样便改善了编码性能。在较高的码率下，必须使用较小的窗口尺寸以便满足这种数据尺寸限制。这必然会使编码性能降低，但对较高码率来说这影响不大。而且，这种将 PCM 数据分成帧的方式给了解码器 18 时间，在全部输出帧读入缓冲器之前能够初始化以启动放音。这样可减少音频编码器的延时或滞后时间。

[0046] 编码器 12 使用了高分辨率滤波器组，优选方法是它可根据码率不同选择不完全（NPR）和完全（PR）重构滤波器，以便把每个音频声道 14 分解为多个子带信号。预测编码器和矢量量化（VQ）编码器分别用来对低频段和高频段子带进行编码。VQ 子带的起始频段或可固定或可根据当前信号的特性来动态确定。在低码率情况下，可以采用联合频率编码同时对多个声道的较高频子带进行编码。

[0047] 预测编码器的优选方法是可根据子带预测增益在 APCM 和 ADPCM 模式之间转换。瞬态分析器把每个子带的子帧分段为预回波和后回波信号（子子帧）并计算预、后回波子子帧各自的比例因数，从而降低预回波失真。编码器根据所有 PCM 声道及当前帧各子带的需要不同（应用心理声学或 mse），从可用的码率中合理地调节分配它们各自的码率，以达到最佳编码效率。通过把预测编码和心理声学模型彼此结合，低码率编码效率得到提高，从而降低了达到主观透明度所需的编码率。诸如计算机或键盘等可编程控制器 19 可用来与编码器 12 相连，以传达输入音频模式信息参数，包括理想码率、声道数量、PR 或 NPR 重构、采

样频率及传输率等。

[0048] 编码信号和辅助信息（即附图中的边信息）被打包并多路复用式地合并为数据流 16, 其形式将解码计算负荷限制在所需的范围内。数据流 16 既可编码到例如 CD、数字视盘 (DVD) 等传输媒体 20 或可通过直播卫星进行传输广播。解码器 18 对各子带信号进行解码并完成逆滤波操作以产生多通道音频信号 22, 该信号主观质量上相当于原始多通道音频信号 14。诸如家庭影院或多媒体计算机等等的音频系统 24 可为用户播放音频信号。

[0049] 多通道编码器

[0050] 如图 2 所示, 编码器 12 包括多个单独的通道编码器 26, 合适选择为五个（左前、中置、右前、左后和右后）, 每通道编码器产生其相应一组编码子带信号 28, 合适选择为每个声道 32 个子带信号。编码器 12 采用全局比特管理 (GBM) 系统 30, 对可用的、共同的比特码率之和（共同比特池）在声道之间、各声道子带之间及各子带的各帧数据之内进行动态比特码率分配。编码器 12 还利用各声道在较高频率子带上可能的相关特性会相应采用联合频率编码技术。此外, 编码器 12 可以在不易感觉到的较高频率子带上使用 VQ 以便在非常低的码率下提供基本的高频保真度或氛围。用这种方式, 编码器利用了不同的信号要求, 例如, 多个声道的子带 rms（均方根）值和心理声学掩蔽等级, 每个声道信号能量随频率的非均匀分布及其在任给定帧内随时间的变化。

[0051] 比特分配概述

[0052] GBM 系统 30 首先确定将对哪些声道的子带进行联合频率编码并对其数据进行平均, 然后确定对哪些子带进行 VQ 编码并从总可用码率中减去其所用码率。对哪些子带进行 VQ 编码可以被认作为首先可定的, 譬如将所有大于某个频率阈值的子带都应用 VQ 编码, 或由每帧中各子带的心理声学掩蔽效应来决定。之后, GBM 系统 30 应用心理声学掩蔽效应对剩余各子带分配其比特率 (ABIT) 从而达到优化解码音频信号主观质量的目的。如果尚有附加比特码率, 编码器可以转换到纯 mmse 方案, 即“注水式”分配 (waterfilling), 并根据子带相应的 rms 值对所有的码率进行再分配从而使误差信号的 rms 值降至最小。这种方法可在极高码率的情况下运用。优选的方法是保留心理声学码率分配的结果, 只是将附加比特码率按照 mmse 方案进行分配。这样可保持用心理声学掩蔽而产生的噪声信号的形状, 并均匀地降低了其噪声基底。

[0053] 另一种方法是对以上优选方法进行修正, 使其附加比特码率根据 rms 值和心理声学掩蔽等级之间的差来进行分配。结果是, 心理声学分配曲线随着码率的不断增加变成成为 mmse 分配曲线, 从而在两种技术之间形成了平滑过渡。上述技术特别适用于固定码率系统。此外, 编码器 12 可以按主观条件或 mse 设定失真等级, 并允许改变总比特码率以保持失真等级。多路复用器 32 按照设定的数据格式把子带信号和辅助信息多路合并为数据流 16。具体的数据格式将在下面的图 20 中讨论。

[0054] 基带编码

[0055] 对于 8-48kHz 范围的采样频率而言, 如图 3 中所示的通道编码器 26 采用了均匀式 512- 节、32- 频带的分析滤波器组 34, 该滤波器组 34 以 48kHz 的采样频率工作, 把每个声道 0-24kHz 的音频谱分解成 32 个带宽为 750Hz 的子带。编码部分 36 对每个子带信号进行编码并将它们多路复用 38 成压缩数据流 16。解码器 18 接收压缩的数据流, 利用拆包器 40 分解出每个子带的编码数据, 对每个子带信号 42 进行解码, 并采用 512- 节、32- 频带的均匀

式内插滤波器组 44 对每个声道重构 PCM 数字音频信号 ($F_{\text{samp}} = 48\text{kHz}$)。

[0056] 在本结构中,所有编码策略,例如 48kHz、96kHz 或 192kHz 的采样频率,都在其最低音频基带上,例如在 0-24kHz,使用了 32- 频带编码 / 解码方法。因此,目前根据 48kHz 采样频率设计和制成的解码器能与将来设计的、利用更高频率分量的编码器兼容。前期存在的解码器总能读取编码信号中的基带部分 (0-24kHz) 并丢弃更高频率编码数据。

[0057] 高采样率编码

[0058] 对于 48-96kHz 范围的采样频率而言,优选法是通道编码器 26 把音频频谱分成两部分,并对下半部分采用均匀式 32- 频带分析滤波器组而对上半部采用 8- 频带的分析滤波器组。如图 4a 和图 4b 所示,0-48kHz 的音频频谱首先采用 256- 节、2- 频带抽取预滤波器组 46 被一分为二,形成的带宽为每频段 24kHz。下半部分频带 (0-24kHz) 如以上按图 3 所述方式分割成 32 个均匀频带并编码。然而上半部分频带 (24-48kHz) 则被分割成 8 个均匀频带进行编码。如果 8- 频带抽取 / 内插滤波器组 48 的延时与 32- 频带滤波器组的相应延时值不相等,则必须在 24-48kHz 信号通路中加入延时补偿 50,以确保在解码器中两个时域波形在进入 2- 频带重组滤波器组之前对齐。在 96kHz 采样频率编码系统中,24-48kHz 音频频带被延迟了 384 个采样点,然后用 128- 节内插滤波器组将其分成 8 个均匀频带。每个 3kHz 频宽的子带被分别进行编码 52,其数据与来自 0-24kHz 频带的编码数据打包 54 形成压缩数据流 16。

[0059] 当到达解码器 18 时,压缩数据流 16 被进行拆包分离 56,用于 32- 频带解码器 (0-24kHz 区) 和 8- 频带解码器 (24-48kHz 区) 的编码数据分别被送到它们各自的解码级 42 和 58 中。其 8 个和 32 个解码子带分别用 128- 节和 512- 节的均匀式内插滤波器组 60 和 44 进行重构。解出的子带然后用 256- 节、2- 频带均匀式内插滤波器组 62 重新结合以产生单一的采样频率为 96kHz 的 PCM 数字音频信号。如果解码器需要以压缩数据流采样频率的一半进行操作,可以方便地通过放弃高端频带编码数据 (24-48kHz) 并只解出 0-24kHz 音频区域内的 32- 音频子带便可达到。

[0060] 通道编码器

[0061] 在上述所有的编码策略中,32- 频带编码 / 解码方法是针对音频带宽为 0-24kHz 间的基带部分进行的。如图 5 所示,取帧器 64 为 PCM 声道开窗将其分段为连续的数据帧 66。PCM 音频窗确定了连续输入的采样数量,该输入采样数量通过编码过程产生数据流中的一幅输出帧。窗口尺寸是根据压缩量即传输率与采样频率之比来设定的,以限制每帧内的编码数据量。每个连续的数据帧 66 通过 32- 频带、512- 节的 FIR (有限冲击响应) 抽取滤波器组 34 被分成 32 个均匀频带 68。每个子带的输出采样数据被缓冲并应用于 32- 频带编码功能级 36。

[0062] 分析级 70 (将在图 10-19 中详细描述) 为被缓冲的子带采样数据产生最佳预测器系数、差分量化器比特率分配和最佳量化器比例因数。分析级 70 还可以在没有预置定值时,决定将对哪个子带进行矢量量化 (VQ) 和对哪些声道进行联合频率编码。这些数据或辅助信息被向前送到选定的 ADPCM 级 72、VQ 级 73 或联合频率编码 (JFC) 级 74 以及数据多路复用器 32 (打包器)。子带采样数据然后通过 ADPCM 或 VQ 方法进行编码,且量化后的编码被输入至多路复用器。JFC 级 74 实际上并不对子带采样数据进行编码,而是产生代码字以指示哪些声道的子带被联合处理以及它们的编码被放于数据流何处。来自每个子带的量化

编码和辅助信息被打包形成数据流 16 并传输送入解码器。

[0063] 到达解码器 18 时,数据流被多路分解 40 或拆包变回成各自的子带数据。比例因数和比特率分配首先被设置装入逆量化器 75 中,每个子带的预测器系数亦被同时装入。然后差分码可直接利用 ADPCM 方法 76 或逆 VQ 方法 77 进行重构,或对指定子带进行逆 JFC 处理 78。最后这些子带用 32- 频带内插滤波器组 44 合并成为单一的 PCM 音频信号 22。

[0064] PCM 信号框构成帧

[0065] 如图 6 所示,当传输率相对于给定采样频率发生变化时,图 5 中所示的取帧器 64 将改变窗口 79 的尺寸,从而使每个输出帧 80 的字节数量被限制在例如 5.3K 字节和 8K 字节之间。表 1 和表 2 分别为对给定采样频率和传输率为设计者提供选择最佳窗口尺寸和解码缓冲器尺寸(帧尺寸)的设计表。在低传输率下,帧尺寸可以相对较大。这使编码器可以利用音频信号在不同时段上的、不平坦的幅值方差分布以改善音频编码器的性能。在高传输率下,需减小帧尺寸以便使字节的总量不会溢出解码缓冲器。结果,设计者在解码器上可使用 8K 字节 RAM 就可以满足所有传输率要求。这降低了解码器的成本。通常,音频窗口的尺寸由下式得出:

[0066]

$$\text{音频窗口} = (\text{帧尺寸}) * F_{\text{samp}} * \left(\frac{8}{T_{\text{rate}}} \right)$$

[0067] 其中帧尺寸是指解码缓冲器的尺寸, F_{samp} 是采样频率,而 T_{rate} 是传输率。音频窗口的尺寸与声道的数量无关。然而,随着声道数量的增加,压缩量也必须相应增加以保持所需的传输率。

[0068]

表 1

[0069]

F_{samp} (kHz)

[0070]	T_{rate}	8-12	16-24	32-48	64-96	128-192
[0071]	$\leq 512\text{kbps}$	1024	2048	4096	★	★
[0072]	$\leq 1024\text{kbps}$	★	1024	2048	★	★
[0073]	$\leq 2048\text{kbps}$	★	★	1024	2048	★
[0074]	$\leq 4096\text{kbps}$	★	★	★	1024	2048

[0075]

表 2

[0076]

F_{samp} (kHz)

[0077]	T_{rate}	8-12	16-24	32-48	64-96	128-192
[0078]	$< 512\text{kbps}$	8-5.3k	8-5.3k	8-5.3k	★	★
[0079]	$< 1024\text{kbps}$	★	8-5.3k	8-5.3k	★	★
[0080]	$< 2048\text{kbps}$	★	★	8-5.3k	8-5.3k	★
[0081]	$< 4096\text{kbps}$	★	★	★	8-5.3k	8-5.3k

[0082] 子带滤波

[0083] 所用的 32- 频带、512- 节均匀式抽取滤波器组 34 是从两种多相滤波器组中选一用来将数据帧 66 分割成图 5 所示的 32 个带宽均匀的子带 68。这两种滤波器组具有不同的重构特性,以重构精度折衷子带编码增益。其中一类滤波器称为完全重构 (PR) 滤波器。当 PR 抽取滤波器(编码)和它相应的内插滤波器(解码)前后直接衔接时,其重构信号是

“完全”的,在此,“完全”被定义为在 24 比特的分辨率时误差小于 0.51sb(最小比特)。另一类滤波器被称为不完全重构(NPR)滤波器,因为其重构信号具有非零的噪声基底值,这与其滤波过程中混叠频率不能完全相抵消的特性有关。

[0084] 图 7 中分别示出了对于单个子带的 NPR 和 PR 滤波器的传递函数 82 和 84。由于 NPR 滤波器不受完全重构要求的限制,它们的邻近阻带排斥(NSBR)比,即通带与第一旁瓣之比,相比 PR 滤波器更大(110dB 相对于 85dB)。如图 8 所示,滤波器的旁瓣使原本处于第三子带中的信号 86 混叠到相邻的子带上。子带增益测量相邻子带中的信号抑制情况,因此表明了滤波器对音频信号的去相关能力。由于 NPR 滤波器比 PR 滤波器具有更大的 NSBR 比,所以它们将具有更大的子带增益。结果,NPR 滤波器提供了更高的编码效率。

[0085] 如图 9 所示,无论是 PR 还是 NPR 滤波器,随着总比特码率的增加,压缩数据流中的总失真将减小。然而,在低码率下,两种滤波器之间子带增益性能的差异大于与 NPR 滤波器相关的噪声底值。因此,NPR 滤波器的相关失真曲线 90 低于 PR 滤波器的相关失真曲线 92。所以,在低码率下音频编码器选择 NPR 滤波器组。当码率增加到某个点 94 时,编码器的量化误差降至 NPR 滤波器的噪声底值之下,继续增加 ADPCM 编码器比特码位不再带来相应的收益。这时候,音频编码器切换使用 PR 滤波器组。

[0086] ADPCM 编码

[0087] ADPCM 编码器 72 根据 H 个先前的重构采样的线性组合产生预测采样 $p(n)$ 。然后从输入 $x(n)$ 中减去该预测采样,从而给出差分采样 $d(n)$ 。这些差分采样数据然后通过用除以 RMS(或 PEAK)比例因数进行比例运算,使比例运算后的差分采样的 RMS 幅值与量化编码器特性曲线 Q 相匹配。比例运算后的差分采样数据 $ud(n)$ 然后被应用到具有 L 级、步长为 SZ 等特征的量化器上,其特征由当前采样分配数据所分配的比特位数 ABIT 决定。量化器对于每个经比例运算的差分采样 $ud(n)$ 产生分级层码 $QL(n)$ 。这些分级层码最终被传送到解码器的 ADPCM 级。为了更新预测器历史,量化器的分级层码 $QL(n)$ 在本地用与量化器 Q 相同特征的逆量化器 $1/Q$ 进行解码以产生量化后、比例运算后的差分采样 $ud'(n)$ 。此数量 $ud'(n)$ 通过与 RMS(或 PEAK)比例因数相乘再进行反比例运算可得到 $d'(n)$ 。通过把起始预测采样 $p(n)$ 与量化差分采样 $d'(n)$ 相加便可重构初始输入采样 $x(n)$ 的量化形式 $x'(n)$ 。然后用该采样更新预测器历史。

[0088] 矢量量化

[0089] 预测器系数和高频子带采样皆使用矢量量化(VQ)进行编码。预测器 VQ 具有 4 个采样值的矢量长度(4 维)和每采样值 3 比特的码率。因此其最后代码本由 4096 个 4 维代码矢量构成。搜索匹配矢量的过程被分成一个两层的树结构,树中的每个节点具有 64 个分枝。顶层存储了 64 个只在编码器中需要用来帮助完成搜索过程的节点代码矢量。底层直接包含了 4096 个在编码器和解码器中都需要的最终代码矢量。就每次搜索而言,需要进行 128 次 4 维 MSE 计算。代码本和顶层的节点矢量是利用 LBG 方法对超过 5 百万个预测系数进行训练优化而成。训练矢量集是通过编制大量音频材料、并在所有显示出明显正向预测增益的子带中累积而成。使用训练集中内的矢量进行测试可得到大约 30dB 的平均 SNRs(信噪比)。

[0090] 高频 VQ 具有 32 个采样数据的矢量长度(32 维,子帧的长度),其码率为每采样值 0.3125 比特。因此最后的代码本由 1024 个 32 维代码矢量构成。匹配矢量的搜索为两层

树结构,树中的每个节点具有 32 个分枝。顶层存储 32 个只在编码器中需要的节点代码矢量。底层包含 1024 个在编码器和解码器中都需要的最终代码矢量。就每次搜索而言,需要进行 64 次 32 维 MSE 计算。代码本和顶层的节点矢量是利用 LBG 方法对超过 7 百万个高频子带采样训练矢量进行训练优化而成。形成训练矢量集的数据是通过编制大量采样频率为 48kHz 的音频材料、并从第 16-32 子带的输出中累积而成。在 48kHz 的采样频率下,这些训练采样代表了 12-24kHz 范围内的音频。使用训练集内的测试矢量预计得到大约 3dB 的平均 SNRs。3dB 的 SNR 尽管很小,但是足以提供高频保真度或在高频上的氛围效应。这在听觉感知上要比简单放弃高频子带的公知技术好得多。

[0091] 联合频率编码

[0092] 在极低比特码率的应用中,总的重构信号保真度可以通过对来自两个或更多通道的高频子带信号之和进行编码以代替逐个独立地编码来加以改善。联合频率编码之所以可行,是因为高频子带往往具有相似的能量分布,而且人的听觉系统主要对高频分量的“强度”而不是对它们的细微结构敏感。于是,因在任何比特码率下都会有更多的比特率用于感知上重要的低段频率的编码,平均来说重构信号提供了很好的综合保真度。

[0093] 联合频率编码索引值 (JOINX) 被直接传送到解码器中以指出哪个通道和子带已经被联合处理以及联合编码信号在数据流中的位置。解码器重构指定声道中的信号并将其复制到其它每个声道中。每个声道然后根据其相应的 RMS 比例因数进行比例运算。由于联合频率编码对时间信号按它们能量分布的相似性进行平均,所以会降低重构保真度。因此它的应用通常限于低比特码率编码应用,而且主要针对 10-20kHz 间的信号。在中高比特码率应用中,联合频率编码通常被停止使用。

[0094] 子带编码器

[0095] 图 10 详细地显示了单个子带运用 ADPCM/APCM 方法的编码过程,特别是描述了如图 5 所示的分析级 70 及 ADPCM 编码器 72 和如图 2 中所示的全局比特管理系统 30 之间的相互作用。图 11-19 详细描述了图 10 中所述的各组件过程。滤波器组 34 把 PCM 音频信号 14 分成 32 个子带信号 $x(n)$, 并写入相应的子带采样缓冲器 96 中。假设音频窗的尺寸为 4096 个采样,每个子带采样缓冲器 96 则存储含有 128 个采样的完整帧,这个帧被分成 4 个 32 采样的子帧。可见,一个含有 1024 采样的窗口尺寸会只产生单一的 32 采样子帧。采样数据 $x(n)$ 被送到分析级 70, 以便确定每个子帧的预测系数、预测模式 (PMODE)、瞬态模式 (TMODE) 和比例因数 (SF)。这些采样数据 $x(n)$ 同时还被提供给 GBM 系统 30, 由系统确定每个声道中每个子带的每个子帧的比特分配 (ABIT)。此后,这些采样数据 $x(n)$ 以每次一个子帧的形式被传递到 ADPCM 编码器 72。

[0096] 最佳预测系数的估计

[0097] 每个子帧的 H 阶 (合适的为四阶) 预测系数可通过对子带采样数据 $x(n)$ 块采用标准自相关方法 98, 即 Weiner-Hopf 或 Yule-Walker 公式, 进行优化处理分别产生。

[0098] 最佳预测系数的量化

[0099] 每组四个预测器系数的量化优选方法是使用如上所述的 4- 元素树搜索、12- 比特矢量代码本 (每个系数 3 比特)。这个 12- 比特矢量代码本包括 4096 个系数矢量, 这些系数矢量已运用标准群集算法并按理想的概率分布特性进行了优化。矢量量化 (VQ) 搜索 100 则选择出一个与最佳系数之间有着最低加权均方差的系数矢量。这些“量化”矢量然后用

来代替每个子帧的最佳系数。一个逆 VQ LUT(查找表)101 用于向 ADPCM 编码器 72 提供量化后的预测器系数。

[0100] 预测差分信号 $d(n)$ 的估计

[0101] 对 ADPCM 来说一个很大的难题是在实现递归过程 72 之前不易预测差分采样序列 $d(n)$ 。对前向自适应子带 ADPCM 的基本要求是在进行 ADPCM 编码之前知道差分信号的能量,以便计算出合适的量化器比特分配,从而明确所产生的量化误差或重构采样信号的噪声量级。差分信号能量的特性也需要了解以便在编码之前确定最佳差分比例因数。

[0102] 遗憾的是,差分信号能量不仅取决于输入信号的特性而且还取决于预测器的性能。除了已知的例如预测器阶数和预测系数优化程度等限制之外,预测器性能还受重构采样信号中引入的量化误差程度或噪声的影响。由于量化噪声是由最终比特分配 ABIT 和差分比例因数 RMS(或 PEAK)值本身来确定,差分信号能量的估计必须通过迭代法来取得 102。

[0103] 步骤 1. 假设量化误差为零

[0104] 差分信号的第一次估计(差分信号估计初值)是将缓冲的子带采样 $x(n)$ 通过 ADPCM 过程而并不对差分信号进行量化。这可以通过在 ADPCM 编码循环中停止量化和 RMS 比例运算功能来实现。用这种方式估计差分信号 $d(n)$,可以从计算中去除比例因数和比特分配值的影响。然而,由于使用了矢量量化的预测系数,这个过程仍然考虑了量化误差对预测器系数的影响。逆 VQ LUT104 被用来提供量化预测系数。为了进一步提高估算预测器的精度,应在计算之前把来自前一数据块处理结束后累积的、真正被 ADPCM 预测器所用的历史采样值复制到当前预测器中。由此可确保预测器可以紧接前一输入缓冲结束时 ADPCM 预测器的真实状态工作。

[0105] 该估计值 $ed(n)$ 和实际过程的 $d(n)$ 之间的主要差别是忽略了量化噪声对重构采样 $x(n)$ 和对降低预测精度的影响。对于有大量层次的量化器来说,通常噪声水平较小(假设通过合适的比例运算),因此实际的差分信号能量与在此估计计算出的结果非常相近。然而,当量化器层次的数量较少时,即在典型的低比特码率音频编码器情况下,实际的预测信号及由此的差分信号能量可能和如此的估计值差别很大。这样便产生了与早先在自适应比特分配过程中所预测值不同的编码噪声底值。

[0106] 尽管如此,预测性能的变化对于所用的应用或比特码率来说影响并不很大。因此,这些估计结果可以不经迭代直接用来计算比特分配和比例因数。另一个改进方法是,如果一个子带很可能被分配使用层数很少的量化器,则可以对差分信号能量有意地作出过高的估计来补偿性能的损失。这种过高估计还可以根据量化器层数的变化进行分级以提高精度。

[0107] 步骤 2. 用估计的比特分配和比例因数进行再计算

[0108] 一旦用差分信号估计初值得出了比特分配 (ABIT) 和比例因数 (SF),就可以将估算出的 ABIT 和 RMS(或 PEAK)值运用于 ADPCM 循环 72,进行进一步的 ADPCM 估计过程以测试它们的最佳性。与估计初值时一样,在开始计算之前,实际的 ADPCM 预测器系数被复制用作估计预测器历史,从而保证两次预测器运算从同一点开始。在缓冲的输入采样数据全部经过第二估计循环处理之后,所得的每个子带中的噪声底值与在自适应比特分配过程中预测的噪声底值进行比较。任何明显的差异则通过修正比特分配和 / 或比例因数进行补偿。

[0109] 步骤 2 可以重复使用以适当改进噪声底值在子带上的分布,每次重复时需用最新

的估计差分信号来计算下一组比特分配和比例因数。通常,如果比例因数的变化大于约 2-3dB,则需重新计算。否则,比特分配可能会违背由心理声学掩蔽过程或 mmse 过程产生的信号-掩蔽比。一般来说,一次重复就足够了。

[0110] 子带预测模式 (PMODE) 的计算

[0111] 为了提高编码效率,预测过程可以被一个控制器 106 在当前子帧中的预测增益降到某一阈值以下时,通过设置 PMODE 指示值任意地终止。当在估计阶段对输入采样块测得的预测增益(输入信号的能量与估计的差分信号能量之比)超出某个正阈值时,PMODE 指示值将置 1。相反,如果测得的预测增益小于正阈值,相应子带的 ADPCM 预测器系数则同时在编码器和解码器中置 0,而且其 PMODE 也置 0。预测增益阈值的设定必须补偿等于因传输预测器系数矢量而使用消耗的比特码所带来的失真率。这样做是为了确保当 PMODE = 1 时,ADPCM 过程的编码增益总是大于或等于前向自适应 PCM(APCM) 编码过程的增益。不然则将 PMODE 置零并将其预测器系数重置为零,ADPCM 过程就简单地转换成 APCM。

[0112] 如果 ADPCM 编码增益的变化对应用来说不是很重要,则可以在任何或所有子带中将 PMODEs 置于高电平。相反,PMODEs 亦可以为以下情况置于低电平,例如某些子带完全不用编码,或是应用的比特率足够高而不需要用预测增益来保持音频的主观质量,或是信号的瞬变含量很高,或者象在音频剪辑应用情况下 ADPCM 编码音频的剪辑接合性不够令人满意等。

[0113] 每个子带相应的预测模式 (PMODE) 值被分离传输且其速率等于编码器和解码器 ADPCM 过程中的线性预测器的更新速率。PMODE 参数的用途是向解码器传送并指示某特定子带是否在其编码音频数据块中含有任何相关的预测器系数矢量地址。当任何子带的 PMODE = 1 时,数据流中将总是包含其预测器系数矢量地址。当任何子带的 PMODE = 0 时,数据流中则不会包含预测器系数矢量地址,编码器和解码器的 ADPCM 级预测器系数必须置 0。

[0114] PMODE 的计算首先对缓冲的子带输入信号能量及其相应缓冲的、从第一阶段估计所得的估计差分信号能量进行对比分析,这里假定没有量化误差。每个子带的输入采样数据 $x(n)$ 和估计的差分采样数据 $ed(n)$ 皆被分开缓冲处理。缓冲器尺寸等于每个预测器更新周期内所包含的采样数,例如子帧的尺寸。预测增益则可按下列公式计算:

$$[0115] \quad P_{\text{gain}}(\text{dB}) = 20.0 * \text{Log}_{10}(\text{RMS}_{x(n)} / \text{RMS}_{ed(n)})$$

[0116] 其中 $\text{RMS}_{x(n)}$ = 缓冲输入采样 $x(n)$ 的均方根值, $\text{RMS}_{ed(n)}$ = 缓冲估计差分采样 $ed(n)$ 的均方根值。

[0117] 正的预测增益表示差分信号平均起来小于输入信号,所以对于相同的比特码率,用 ADPCM 过程相比 APCM 可以降低重构信号的噪声底值。负增益则表示 ADPCM 编码器产生的差分信号平均大于输入信号,这导致了在相同比特码率下比 APCM 更高的噪声底值。通常,用以启用 PMODE (即置 1) 的预测增益阈值是正的,而且其值已经考虑了因传输预测器系数矢量地址而消耗的额外通道容量。

[0118] 子带瞬态变化模式 (TMODE) 的计算

[0119] 控制器 106 计算每个子带中每个子帧的瞬态模式 (TMODE)。TMODEs 指示了比例因数和采样数据的数量及其相应有效的部分,这些采样数据当 PMODE = 1 时是缓冲器中的估计差分信号 $ed(n)$,当 PMODE = 0 时是缓冲器中的输入子带信号 $x(n)$ 。TMODEs 被传送到解

码器,其更新频率与预测系数矢量地址相同。瞬态模式的目的是在信号出现瞬变时降低编码带来的、可听见的“预回波”人为噪音。

[0120] 瞬变可定义为在低幅值信号和高幅值信号之间的快速过渡。由于比例因数在一整块子带差分采样数据上进行平均运算而来,那么如果在这信号块上发生快速幅值变化,即瞬变,所计算出的比例因数往往比那些处于瞬变前的低幅值采样需要的最佳值大得多。因此对于瞬变前的采样数据来说,量化误差可能会很大。这种噪声听觉上则被称为预回波失真。

[0121] 在实际运用中,瞬态模式是用来修正用于平均计算子带比例因数的数据块长度,以限制瞬变对那些紧靠瞬变前的差分采样的比例运算的影响。这样做的动机是因为人听觉系统中存在的、固有的预掩蔽现象,此现象表明在出现瞬态变化时,它之前的噪声如果持续时间很短则可被瞬态变化本身掩蔽不被察觉。

[0122] 根据 PMODE 的取值不同,子带采样缓冲器中的 $x(n)$ 的内容,即子帧,或估计的差分缓冲器 $ed(n)$ 的内容被复制到瞬变分析缓冲器中。根据分析缓冲器的采样尺寸,缓冲器中的内容均匀地分成 2、3 或 4 个子子帧。例如,如果分析缓冲器包含了 32 个子带采样 ($21.3\text{ms}@1500\text{Hz}$),缓冲器可分区为各含有 8 个采样的 4 个子子帧,在子带采样率为 1500Hz 的情况下,其时间分辨率为 5.3ms 。换一种情形,如果分析窗由 16 个子带采样构成,那么只需将缓冲器分成两个子子帧以提供相同时间分辨率。

[0123] 每个子子帧内的信号被加以分析并确定除第一子子帧之外的、每个子子帧的瞬变模式状态。如果任何子子帧被认作为瞬态,则将为分析缓冲器即当前子帧产生两个独立的比例因数。第一个比例因数是从瞬变子子帧之前子子帧中的采样计算而来。第二个比例因数则根据瞬变子子帧中的采样并结合所有以后子子帧计算而来。

[0124] 第一子子帧的瞬变状态不用计算,因为其处于分析窗开始的位置本身已可自动限制其量化噪声。如果有一个以上的子子帧被认作出现瞬变,则只考虑首先出现的那个子子帧。如果没有子缓冲器被检测到出现瞬变,则只需用分析缓冲器中的所有采样数据计算单个比例因数。用这种方式,用瞬变采样数据计算的比例因数值不会用于超出一子子帧时段之前的早期采样数据的比例运算。由此,把预瞬变量化噪声限制在一个子子帧周期之内。

[0125] 瞬态的确认声明

[0126] 如果一子子帧相比前一子缓冲器的能量比超出瞬变阈值 (TT),而且前一子子帧中的能量低于预瞬变阈值 (PTT) 则声明在该子子帧中有瞬变。TT 和 PTT 的取值决定于比特码率和所需的预回波抑制程度。这些值通常可以变化调节直至感到的预回波失真与其它人为编码噪声 (如果有的话) 能级相近。增加 TT 和 / 或减小 PTT 的值都将减小子子帧被认作为含有瞬变的可能性,由此降低了用于比例因数传输的比特码率。反之,减小 TT 和 / 或增加 PTT 的值将增加子子帧被认作为含有瞬变的可能性,并由此增加了用于比例因数传输的比特码率。

[0127] 由于 TT 和 PTT 是针对每个子带分别设定的,所以编码器中所有子带的瞬态检测的灵敏度可以自由设定。例如,如果发现高频子带中的预回波相比低频子带中的预回波不易觉察,那么其阈值可以相应设定来减小高频子带被认作为含有瞬变的机会。此外,由于 TMODEs 被嵌入压缩的数据流中,解码器不必知道在编码器中使用的瞬变检测算法亦可对 TMODE 信息进行适当的解码。

[0128] 四种子缓冲器的结构配置

[0129] 如图 11a 所示,如果子带分析缓冲器 109 中的第一子子帧 108 出现瞬变,或如果没有检测到任何瞬变子子帧,则 $TMODE = 0$ 。如果第二子子帧出现瞬变而第一子子帧没有,则 $TMODE = 1$ 。如果第三子子帧出现瞬变而第一或第二子子帧都没有,则 $TMODE = 2$ 。如果只有第四子子帧出现瞬变则 $TMODE = 3$ 。

[0130] 比例因数的计算

[0131] 如图 11b 所示,当 $TMODE = 0$ 时,比例因数 110 在所有子子帧上计算。当 $TMODE = 1$ 时,第一个比例因数在第一子子帧上计算,第二个比例因数在所有之后的子子帧上计算。当 $TMODE = 2$ 时,第一个比例因数在第一和第二子子帧上计算,第二个比例因数在所有之后的子子帧上计算。当 $TMODE = 3$ 时,第一个比例因数在第一、第二和第三子子帧上计算,第二个比例因数在第四子子帧上计算。

[0132] 用 $TMODE$ 进行 ADPCM 编码和解码

[0133] 当 $TMODE = 0$ 时,整个分析缓冲器期间即子帧内的子带差分采样数据用单个比例因数进行比例运算,该比例因数还传至解码器以进行反比例运算。当 $TMODE > 0$ 时,需要两个比例因数对子带差分采样数据进行比例运算且两者皆需传至解码器。不管是什么 $TMODE$,在一组差分采样数据上产生的比例因数只用于这组数据的比例运算。

[0134] 子带比例因数 (RMS 或 PEAK) 的计算

[0135] 根据每个子带的 $PMODE$ 取值不同,用来计算其比例因数的数据或是估计的差分采样 $ed(n)$,或是输入的子带采样 $x(n)$ 。 $TMODEs$ 在此计算中则用来确定比例因数的数量以及在缓冲器中它们对应的子子帧。

[0136] RMS 比例因数计算

[0137] 对第 j 个子带而言,rms 比例因数可按下列公式计算:

[0138] 当 $TMODE = 0$ 时,单个 rms 的值是:

$$[0139] \quad RMS_j = \left(\sum_{n=1}^L ed(n)^2 / L \right)^{0.5}$$

[0140] 其中 L 是子帧中的采样数量。

[0141] 当 $TMODE > 0$ 时,那么两个 rms 值为:

$$[0142] \quad RMS1_j = \left(\sum_{n=1}^k ed(n)^2 / L \right)^{0.5}$$

$$[0143] \quad RMS2_j = \left(\sum_{n=k+1}^{k+1} ed(n)^2 / L \right)^{0.5}$$

[0144] 其中 $k = (TMODE * L / NSB)$, NSB 是均匀尺寸子子帧的数量。

[0145] 如果 $PMODE = 0$,则用输入采样 $x_j(n)$ 替换差分采样 $ed_j(n)$ 。

[0146] PEAK 比例因数的计算

[0147] 就第 j 个子带而言,峰值比例因数可按下列公式计算:

[0148] 当 $TMODE = 0$ 时,单个峰值是:

[0149] $PEAK_j = \text{MAX}(\text{ABS}(ed_j(n)))$, $n = 1, L$

- [0150] 当 $TMODE > 0$ 时,两个峰值是:
- [0151] $PEAK1_j = \text{MAX}(\text{ABS}(ed_j(n))), n = 1, (TMODE * L/NSB)$
- [0152] $PEAK2_j = \text{MAX}(\text{ABS}(ed_j(n))), n = (1+TMODE * L/NSB), L$
- [0153] 如果 $PMODE = 0$,则用输入采样 $x_j(n)$ 替换差分采样 $ed_j(n)$ 。
- [0154] $PMODE$ 、 $TMODE$ 和比例因数的量化
- [0155] $PMODE_s$ 的量化
- [0156] 预测模式标记值只取两个值,开或关,可直接作为 1- 比特编码送至解码器。
- [0157] $TMODE_s$ 的量化
- [0158] 瞬变模式标记值最多具有 4 个值:0、1、2 和 3,它可直接作为 2- 比特无符号整数代码送至解码器,或可通过使用一个 4- 层熵编码表以争取将传输 $TMODE_s$ 的平均字长降至 2 比特以下。通常,熵编码只是在低比特码率应用时才选择运用以节省比特位数。
- [0159] 图 12 中详细示出的熵编码过程 112 可描述如下:把 j 个子带的瞬态模式码 $TMODE(j)$ 与多个 (p) 4- 层中值上升、可变长度代码本进行相配比较,其中每个代码本是根据不同的输入统计特性优化设计的。 $TMODE$ 值与这些 4- 层表 114 进行相配比较并计算与每个表相关的总比特位数用量 (NBp) 116。在相配过程中能提供最少比特用量的码表就被选中并记作 THUFF 索引值。从该表格中取出的匹配码字 $VTMODE(j)$,将其与 THUFF 索引字一道被打包并送到解码器。拥有同样一组 4- 层反向表格的解码器可利用 THUFF 索引值将输入的可变长度编码 $VTMODE(j)$ 送至合适的表格并解出 $TMODE$ 索引值。
- [0160] 子带比例因数的量化
- [0161] 为了将比例因数传送到解码器必须将它们量化成已知的编码格式。在该系统中,比例因数通过使用均匀 64- 层对数特性、或均匀 128- 层对数特性、或可变速编码的均匀 64- 层对数特性量化器进行量化 120。其中,两种 64- 层的量化器显示的步长皆为 2.25dB,而 128- 层的步长为 1.25dB。64- 层量化用于低至中比特码率,附加的可变速率编码用于低比特码率应用,而 128- 层通常用于高比特码率应用。
- [0162] 图 13 示出了量化过程 120。以 RMS 或 PEAK 表示的比例因数先从缓冲器 121 中读出,转换成对数域 122,然后根据编码器模式控制器 128 的判断将其送到 64- 层或 128- 层均匀量化器 124、126。然后把对数量化的比例因数写入缓冲器 130 中。128- 层和 64- 层量化器的范围分别可以满足动态范围约为 160dB 和 144dB 的比例因数。把 128- 层的上限设定为能覆盖 24- 位输入 PCM 数字音频信号的动态范围。把 64- 层的上限设定为能覆盖 20- 位输入 PCM 数字音频信号的动态范围。
- [0163] 对数比例因数然后与量化器进行相配比较并用最接近的量化器层代码 RMS_{ql} (或 $PEAK_{ql}$) 代替比例因数。在使用 64- 层量化器的情况下,这些代码为 6- 比特长,代码范围为 0-63。在使用 128- 层量化器的情况下,代码长度为 7- 位,其范围为 0-127。
- [0164] 逆量化 131 可简便地通过把各层代码运用其各自的逆量化特性实现,以产生 RMS_q (或 $PEAK_q$) 值。对 ADPCM (或当 $PMODE = 0$ 时为 APCM) 差分采样的比例运算而言,编码器及解码器都使用了量化过的比例因数,由此可确保比例运算和反比例运算在两处的过程同步一致。
- [0165] 如果 64- 层量化器编码的比特码率仍需要降低,则要进一步进行熵、或变字长编码。 j 个子带的 64- 层编码从第二子带 ($j = 2$) 开始到最高有效子带被进行一阶差分编

码 132。该过程也可以用于对 PEAK 比例因数进行编码。有符号的差分编码 $DRMS_{ql}(j)$ (或 $DPEAK_{ql}(j)$) 的最大范围为 ± 63 而且将这些编码存储在缓冲器 134 中。为了在原始的 6-位编码上降低它们的比特码率,这些差分编码与多个 (p) 127- 层中值上升、可变长度代码本进行相配比较,其中每个代码本是根据不同的输入统计特性优化设计的。

[0166] 对有符号的差分编码进行熵编码的过程与图 12 中所示用于瞬变模式的熵编码过程相同,只是使用了 p 个 127- 层可变长度代码本。在比较过程中提供最低比特用量的表格则以 SHUFF 索引值形式被选中。其匹配的编码 $VDRMS_{ql}(j)$ 从表格中取出,与 SHUFF 索引字一道打包并传送到解码器。具有同样一组 (p) 127- 层反向表的解码器,可利用 SHUFF 索引值把进入的可变长度编码送入合适的表中,以便对其进行解码使之回到差分量化器代码层。下列程序可用来将差分代码层转回成绝对值:

[0167] $RMS_{ql}(1) = DRMS_{ql}(1)$

[0168] $RMS_{ql}(j) = DRMS_{ql}(j) + RMS_{ql}(j-1) \quad j = 2, \dots, K$

[0169] 而用下列程序可使 PEAK 差分代码层转回成绝对值:

[0170] $PEAK_{ql}(1) = DPEAK_{ql}(1)$

[0171] $PEAK_{ql}(j) = DPEAK_{ql}(j) + PEAK_{ql}(j-1) \quad j = 2, \dots, K$

[0172] 在以上两种情况下, K = 有效子带的数量。

[0173] 全局比特分配

[0174] 图 10 所示的全局码率管理系统 30 在多通道音频编码器中管理比特分配 (ABIT), 确定有效子带 (SUBS) 的数量和联合频率策略 (JOINX) 以及 VQ 策略, 以在降低的比特码率情况下提供主观透明编码。这不但增加了固定媒体上可编码和储存的音频声道的数量并 / 或延长播放时间,同时还保持或提高了音频保真度。通常, GBM 系统 30 首先根据编码器中经预测增益修正的心理声学分析结果将比特分配到每个子带。然后根据 mmse 方案分配所剩余的比特以便降低总噪声底值。为了优化编码效率, GBM 系统同时考虑所有声道、所有子带及整个数据帧并进行比特分配。此外, 可以利用联合频率编码策略。用这种方式, 系统充分利用了声道之间、频率范围内及时域上的信号能量的非均匀分布特性。

[0175] 心理声学分析

[0176] 心理声学测量是用来决定音频信号中存在的、感知上不相关的信息。感知上不相关的信息可定义为音频信号中不能为人类听众所听到的部分, 它可以在时域上、频域上或以其它一些方式进行测量。J. D. 约翰斯顿 (J. D. Johnston): “采用感知噪声标准的音频信号变换编码”, 见 IEEE Journal on Selected Areas in Communications, 第 JSAC-6 期, 第 2 号, 第 314 ~ 323 页, 1988 年 2 月, 其中描述了心理声学编码的总则原理。

[0177] 两个主要因素将影响心理声学测量。一个是人类听力的、与频率有关的绝对阈值。另一个是掩蔽效应, 即人听到的第一声音能将其同时演奏或甚至在其之后的第二声音掩盖住的事实。换句话说, 第一声音能阻止我们听到第二声音, 也就是说将其掩蔽掉。

[0178] 在子带编码器中, 心理声学计算的最终结果是一组数, 指定在某瞬间对于每个子带的、不再能听觉的噪声量级。该计算方法是公知的而且被并入 MPEG1 压缩标准 ISO/IEC DIS 11172 “信息技术 - 用于约 1.5Mbits/s 以内的数字存储媒体的运动图象和相关声音的编码” 1992。这些数随音频信号动态变化。编码器借助比特分配过程以调节子带中的量化噪声底值, 以使这些子带中的量化噪声小于可听觉的量级。

[0179] 精确的心理声学计算通常需要在时间-频率变换中具备高频率分辨率。这意味着时间-频率变换时需要较大的分析窗。标准的分析窗尺寸是 1024 个采样,相应于压缩音频数据的子帧。长度为 1024 的 fft 的频率分辨率大致与人耳的时间分辨率匹配。

[0180] 心理声学模型的输出对 32 个子带中每一个都产生了一个信号-掩蔽 (SMR) 比。SMR 表示了其子带所能承受的量化噪声量,因此还表示了量化其子带采样数据所需的比特位数。具体地说,大的 SMR ($\gg 1$) 表示需要的比特位数很多,而小的 SMR (> 0) 则表示需要的比特位数较少。如果 $SMR < 0$,则音频信号低于噪声掩蔽阈值,这时不需要量化比特。

[0181] 如图 14 所示,每个连续帧的 SMR 通常通过下列步骤产生。1) 对 PCM 音频采样数据进行 fft 计算,优选长度为 1024,得出一系列频率系数 142,2) 对每个子带将产生的频率系数与其心理声学、频率相关的音调和噪声掩蔽值 144 进行卷积,3) 对每个子带上产生的系数进行平均以得出 SMR 的量级,和 4) 作为可选步骤,根据图 15 所示的人的听觉响应 146 对 SMRs 进行归一化处理。

[0182] 人耳的灵敏度在频率接近 4kHz 时最高并随着频率的进一步升高或降低而下降。因此,要想感受相同的音量强度,20kHz 的信号必须比 4kHz 的信号强得多。因此一般来说,4kHz 频率周围的 SMRs 与边远频率比较要重要得多。然而,曲线的精确形状与传送给听者的信号平均功率有关。随着音量的增加,听觉响应范围 146 受到压缩。因此,在某特定音量之下优化的系统对其它音量来说只是次优的。结果是,或者选择一指定功率级对 SMR 量级进行归一化处理,或者不用归一化处理。图 16 中示出了产生的用于 32 个子带的 SMRs 148。

[0183] 比特分配程序

[0184] GBM 系统 30 首先选择合适的编码策略,决定哪些子带用 VQ 和 ADPCM 算法进行编码以及是否启用 JFC。之后,GBM 系统将选择心理声学或者是 MMSE 比特分配方法。例如,在高比特码率下,系统可能停用心理声学模式而使用真正的 mmse 分配方案。这样可降低计算的复杂性而且在重构音频信号中并不感到有任何听觉变化。相反,在低速率下,系统能够启用如上所述的联合频率编码方案从而提高较低频率的重构保真度。GBM 系统能够根据信号中的瞬变含量在帧与帧之间进行正常心理声学分配和 mmse 分配方法的切换。当瞬变含量高时,计算 SMRs 时使用的稳态假设就不再有效,因此 mmse 方案可以提供更好的特性。

[0185] 就心理声学分配方法而言,GBM 系统首先分配可用比特以满足心理声学效应,然后对剩余比特进行分配以便降低总噪声底值。第一步如上所述是确定当前帧每个子带的 SMRs。下一步是按各子带中的预测增益 (Pgain) 调节其 SMRs 从而产生掩蔽-噪声比 (MNRs)。其原理是 ADPCM 编码器将提供一部分所需的 SMR。所以用更少的比特位数便可达到听不到的心理声学噪声级。

[0186] 假设 $PMODE = 1$,则第 j 个子带的 MNR 由下式给出:

[0187] $MNR(j) = SMR(j) - Pgain(j) * PEF(ABIT)$

[0188] 其中 $PEF(ABIT)$ 是量化器的预测效率因数。为了计算 $MNR(j)$,设计者必须估计比特分配 (ABIT) 情况,这可以通过只用 $SMR(j)$ 比值进行比特分配或通过假设 $PEF(ABIT) = 1$ 而得到。在中高比特码率下,有效预测增益近似等于计算的预测增益。然而,在低比特码率下,有效预测增益将降低。用例如 5-层量化器得到的有效预测增益近似为估计的预测增益的 0.7 倍,而 65-层量化器则使有效预测增益近似等于估计的预测增益, $PEF = 1.0$ 。在极限情况下,当比特码率为零时,实际上预测编码被停止使用,有效预测增益为零。

[0189] 在下一步骤中,GBM 系统 30 产生一个满足每个子带 MNR 的比特码位分配方案。这是利用 1 比特约等于 6dB 的信号失真的近似方法而实现的。为了确保编码失真小于心理声学听觉阈值,分配的比特率是 MNR 除以 6dB 得到值上舍入后的最大整数,由下式给出:

[0190]

$$ABIT(j) = \left\lceil \frac{MNR(j)}{6dB} \right\rceil$$

[0191] 通过用这种方式进行比特分配,重构信号中的噪声级 156 将如图 17 所示随信号本身 157 而变化。因此,在信号很强的频率上,噪声级将比较高,但仍将保持在听力察觉范围之外。在信号比较弱的频率上,噪声底值将很小且不会被听到。使用这种心理声学模型的平均误差总是大于 mmse 噪声级 158,但是就其听力可感知部分来说性能更好,特别是在低比特码率下更是如此。

[0192] 在所有声道、每个子带上分配的比特总和大于或小于目标比特码率的情况下,GBM 程序将重复迭代以减小或增加各子带的比特分配。另外一种方法是计算每个声道的目标比特码率。这虽然是次佳方法但硬件实现时特别容易。例如,可用比特可以在声道中均匀地分布,或是按每个声道的平均 SMR 或 RMS 成比例分布。

[0193] 在局部比特分配总和(包含 VQ 码位和辅助信息在内)超过目标比特码率的情况下,全局码率管理程序将逐步降低局部子带的比特分配。有多种具体方法可用于降低平均比特码率。首先,用于计算比特码率的上舍入整数函数可变为下舍入整数函数。其次可以从最小 MNRs 的子带中减去 1 比特。此外,可以停止对较高频率子带的编码或启用联合频率编码。所有降低比特码率的策略均遵循适度地、逐渐地降低编码分辨率的基本原则,首先使用的是感知音质损失最小的策略,损失大的策略则最后使用。

[0194] 在目标比特码率大于局部比特分配总和(包含 VQ 码位和辅助信息在内)的情况下,全局码率管理程序将逐步、迭代地增加局部子带的比特分配以降低重构信号的总噪声底值。这样情况下,先前被分配了零比特的子带可能又进入编码之列。计算这种‘接通’的子带的比特用量时,需要考虑到其可能启用 PMODE 时用于传输任何预测器系数的成本。

[0195] GBM 程序可以从三个不同方案中选择其一以便分配剩余比特。一种方案是对所有比特用 mmse 方法进行再分配以产生近似平坦的噪声底值。这相当于放弃了起初的心理声学模型。为了达到 mmse 噪声底值,图 18a 所示的子带 RMS 值曲线 160 被倒置成如图 18b 所示的形式,所有比特然后进行“注水式”分配直至耗尽。这种公知技术被称为注水是因为随着分配比特位数量的增加失真度均匀地降低。在图中所示的例子中,第一比特分配给了子带 1,第二和第三比特分配给了子带 1 和 2,第四至第七比特分配给了子带 1、2、4 和 7,等等。另外一种方法是先给每个子带分配 1 比特以保证对每个子带进行编码,此后将剩余比特以注水式分配。

[0196] 第二种也是优选的方案是按照上述 mmse 方法和 RMS 曲线来分配剩余比特。这种方法的效果是既均匀降低了如图 17 所示的噪声底值 157 又保持了原心理声学掩蔽曲线形状。这提供了一种很好的介于心理声学 and mse 失真之间的折衷方案。

[0197] 第三种方法是根据子带的 RMS 和 MNR 值之间的差值曲线使用 mmse 方法分配剩余比特。这种方法的效果是,随着比特码率的增加,噪声底值的形状可从最佳心理声学形状

157 平滑过渡至最佳（平坦的）mmse 形状 158。无论使用这些方案的哪一种，如果任何子带中的编码误差相对于源 PCM 来说下降至 0.5LSB 之下，该子带就不再有进一步比特分配。一种可选择的方法是使用固定的子带比特分配最大值来限定各子带可以分配到的最大比特位数。

[0198] 在上面讨论的编码系统中，我们假定了每个采样值的平均比特率是固定的，而且是以重构音频信号保真度最大为目的产生了比特分配。另外一种方法是先固定设置 mse 或感知失真度，然后允许比特码率变化以满足失真度。在 mmse 方法中，RMS 曲线可以被简单地进行注水式分配直至满足失真度量级。所需的比特码率将根据子带的 RMS 量级而改变。在心理声学方法中，则以满足各 MNRs 来进行比特分配。结果是其比特率将根据各 SMRs 和预测增益而改变。这种分配方法目前用途不广，因为当前的解码器均以固定的码率工作。然而，其它的传媒系统如 ATM 或随机存取储存媒体可能在不久的将来使可变码率编码成为实际可用的方法。

[0199] 比特分配索引值 (ABIT) 的量化

[0200] 在全局比特管理过程中，自适应比特分配程序对每个子带和每个声道产生其比特分配索引值 (ABIT)。编码器产生此索引值的目的是为了指示如图 10 所示的必需的量化层数目 162，用此层数在量化差分信号时使解码音频达到主观最佳重构噪声底值。在解码器中，这些索引值指示了逆量化所需的层数。每个分析缓冲窗产生一组索引值，其取值范围为 0-27。索引值、量化层的数目和其相应的差分子带信噪比 SN_{QR} 近似值之间的关系如表 3 所示。由于差分信号被归一化，步长 164 被设定等于 1。

[0201] 表 3

[0202]	ABIT 变址	量化层的数目	编码长度 (位)	$SN_{QR}(\text{dB})$
[0203]	0	0	0	-
[0204]	1	3	可变的	8
[0205]	2	5	可变的	12
[0206]	3	7 (或 8)	可变的 (或 3)	16
[0207]	4	9	可变的	19
[0208]	5	13	可变的	21
[0209]	6	17 (或 16)	可变的 (或 4)	24
[0210]	7	25	可变的	27
[0211]	8	33 (或 32)	可变的 (或 5)	30
[0212]	9	65 (或 64)	可变的 (或 6)	36
[0213]	10	129 (或 128)	可变的 (或 7)	42
[0214]	11	256	8	48
[0215]	12	512	9	54
[0216]	13	1024	10	60
[0217]	14	2048	11	66
[0218]	15	4096	12	72
[0219]	16	8192	13	78
[0220]	17	16384	14	84

[0221]	18	32768	15	90
[0222]	19	65536	16	96
[0223]	20	131072	17	102
[0224]	21	262144	18	108
[0225]	22	524268	19	114
[0226]	23	1048576	20	120
[0227]	24	2097152	21	126
[0228]	25	4194304	22	132
[0229]	26	8388608	23	138
[0230]	27	16777216	24	144

[0231] 比特分配索引值 (ABIT) 可以用 4- 比特无符号的整数代码字、5- 位无符号的整数代码字或用 12- 层熵表直接传输到解码器。通常, 熵编码可用于低比特码率应用以节省比特。ABIT 的编码方法在编码器中通过模式控制来设定并传至解码器。熵编码过程如图 12 所示, 把 ABIT 索引值相配比较 166 到一个由 BHUFF 索引值指定的代码本上, 并从一具有 12- 层 ABIT 表的代码本中匹配比较出特定码 VABIT。

[0232] 全局比特码率控制

[0233] 由于辅助信息和差分子带采样数据皆可以选择用熵编码可变长度代码本进行编码, 所以当以固定的速率传输压缩比特流时, 必须使用某种机制来调节编码器产生的比特码率。由于辅助信息一旦计算后通常不希望更改, 比特码率的调节最好通过重复迭代地改变 ADPCM 编码器中的差分子带采样量化过程来达到, 直至满足比特码率限制条件。

[0234] 在上述系统中, 图 10 中的全局码率控制 (GRC) 系统 178 通过改变分层码值的统计分布来调节在量化器层码与熵表比较相配过程中产生的比特码率。所有熵表都被假定具有层码值越大码字越长的相似趋向。在这种情况下, 平均比特码率随着低值编码层概率的增加而降低, 反之亦然。在 ADPCM (或 APCM) 量化过程中, 比例因数的大小决定了分层编码值的分布或使用。例如, 随着比例因数尺寸的增加差分采样值将趋于在较低层上量化, 因此代码值将逐渐变小。这样又将导致较小的熵代码字长和更低的比特码率。

[0235] 这种方法的缺点是比例因数尺寸的增加相应地、成比例地提升了子带采样中的重构噪声。然而在实际应用中, 对比例因数的调节通常不大于 1dB-3dB。如果需要进行更大的调节, 则最好是回到比特分配以减少总比特分配, 而不应冒险因使用过大的比例因数而使子带中可能出现可听觉的量化噪声。

[0236] 为了调节熵编码 ADPCM 的比特分配, 每个子带的预测历史采样值应存储在临时缓冲器中, 以便 ADPCM 编码过程需要重复的情形。接着, 利用从子带 LPC 分析推导出的预测系数 A_n 以及比例因数 RMS (或 PEAK)、量化器比特分配 ABIT、瞬变模式 TMODE、和从估计的差分信号中推导出的预测模式 PMODE, 所有子带采样缓冲器可通过完整的 ADPCM 过程进行编码。所产生的量化器层代码被进行缓冲并映射到具有最低比特用量的熵可变长度代码本上, 代码本大小再次使用比特分配索引值来确定。

[0237] 随后, GRC 系统对所有索引值分组分析, 对有相同比特分配索引值的每个子带统一计算使用的比特位数。例如, 当 $ABIT = 1$ 时, 全局比特管理中的比特分配计算可以假定每个子带采样为 1.4 的平均比特码率 (即, 最佳层代码幅值分布假设时熵编码本的平均码率)。

如果所有 $ABIT = 1$ 的子带的总比特位数用量大于 $1.4x$ (子带采样的总数), 那么所有这些子带的比例因数都可以增加从而造成比特码率的下降。调节子带比例因数的决定最好是留在得到所有 $ABIT$ 索引值码率之后。由此, 低于比特分配过程中假定码率的索引值可用来补偿那些高于假定比特码率的索引值。这种评估过程可在适当时扩展应用于所有音频声道。

[0238] 为了降低总比特码率, 建议的程序是从超过阈值的最低 $ABIT$ 索引值比特率开始, 增加每个具有这种比特分配率的子带的比例因数。实际达到的比特位数的减少量是这些子带原先高于该分配率指定的码率。如果修正后的比特用量仍然超出允许的最大值, 那么下一个比特用量超出指定值的、较高 $ABIT$ 索引值子带中的比例因数将会增加。这个过程持续进行直到修正的比特用量低于最大值为止。

[0239] 一旦达到这一目的, 就将旧的历史数据装入预测器中, 并且对已经修正了比例因数的那些子带重复进行 ADPCM 编码过程 72。此后, 将层代码再次映射至最佳熵代码本并重新计算比特用量。如果任何一个比特用量仍超过指定码率, 那么要进一步增加比例因数并且重复进行上述循环。

[0240] 修正比例因数的方式有两种。第一种是向解码器对每个 $ABIT$ 索引值发送一个调整系数。例如, 一个 2- 比特字能够代表 0、1、2 和 3dB 的调整范围。由于使用相同 $ABIT$ 索引值的子带都使用相同的调整系数, 而且只有索引值为 1-10 可以使用熵编码, 对于所有子带来说需要传送的调整系数的最大数目是 10。另一种方法是, 通过选择高量化器层可以改变每个子带中的比例因数。然而, 由于比例因数量化器的步长分别为 1.25 和 2.5dB, 所以其比例因数的调节只限于这些步长。此外, 当使用该技术时, 如果启用熵编码的话, 则需要重新计算比例因数的差分编码及其产生的比特用量。

[0241] 一般说来, 当比特率低于所需的码率时, 可以使用同样的程序来增加比特码率。在这种情况下, 比例因数将被减小使得差分采样更好地利用量化器的外高层, 并因此而使用熵表中更长的代码字。

[0242] 如果在合理的迭代运算次数之后, 比特分配索引值的比特用量不能再降低, 或在传送比例因数调节系数时调节步长已经达到极限, 那么有两种可能的修正方法。首先, 可以增加那些码率已在指定范围内的子带的比例因数以此降低总比特码率。另一种方法是, 放弃整个 ADPCM 编码过程并重新计算所有子带的自适应比特分配, 这一次使用更少的比特位数。

[0243] 数据流格式

[0244] 图 10 所示的多路复用器 32 将每个声道的数据打包, 然后把每个声道的打包数据多路复用成输出帧以形成数据流 16。打包和多路复用数据的方法, 即图 19 所示的帧格式 186, 其设计具有下列特性, 使得音频编码器可以用于广大范围的应用之中, 能扩展到更高的采样频率, 每帧内的数据量受到限制, 播放可以在每个子子帧中独立启动从而减少延迟, 以及降低解码错误。

[0245] 如图所示, 单个帧 186 (4096 个 PCM 采样值 / 声道) 确立了比特位数据流的边界, 它含有足够信息以适当解码出相应音频块, 该单个帧由 4 个子帧 188 (1024 个 PCM 采样值 / 声道) 构成, 而每个子帧又由 4 个子子帧 190 (256 个 PCM 采样值 / 声道) 构成。帧同步字 192 位于每个音频帧的开始。帧头信息 194 主要提供与帧 186 的结构、编码器产生比特流时的配置以及各种可选操作特性如嵌入动态范围控制和时间码等有关的信息。可选性的帧头

信息 196 告知解码器是否需要下向声道混音、是否进行了动态范围补偿和数据流是否包含了辅助数据字节。音频编码头信息 198 指示编码器所用的、组装编码‘辅助信息’的打包装置和编码格式,‘辅助信息’即,比特分配、比例因数、PMODES、TMODES、代码本等等。剩余的帧由 SUBFS 个连续音频子帧 188 构成。

[0246] 每个子帧的开始包含了音频编码辅助信息 200,该信息将压缩音频多个关键编码系统的相关信息传至解码器。这些信息包括了瞬变检测、预测编码、自适应比特分配、高频矢量量化、强度编码和自适应比例运算。这些数据中的很多是从数据流中用以上的音频编码头信息拆包取得的。高频 VQ 编码数据列 202 是由用 VQSUB 索引值指示的、每高频子带为 10-比特的索引值组成。低频效应数据列 204 为可选项,代表了用于驱动如低音炮扬声器的极低频率数据。

[0247] 音频数据列 206 用霍夫曼 / 固定逆量化器进行解码,并被分成多个子子帧 (SSC),每子子帧为每个声道可解出最多是 256 个 PCM 采样。超采样频音频数据列 208 只有当采样频率大于 48kHz 时才会存在。为了保持兼容,不能在采样频率高于 48kHz 时工作的解码器应当跳过该音频数据阵列。DSYNC210 是用来验证音频帧中子帧的结束位置。如果此位置没有验证,则该子帧中解出的音频应认作为不可靠。相应结果要么对此帧进行静音处理要么重复前一帧。

[0248] 子带解码器

[0249] 图 20 是子带采样解码器 18 的方框图。解码器与编码器相比相当简单,而且并不涉及对重构音频质量来说具有基本重要意义的计算(例如比特分配)。在同步后,拆包器 40 对压缩的音频数据流 16 进行拆包,检测并按需纠正因传输而引入的错误,并且把数据多路分解为各音频声道。子带差分信号被重新量化成 PCM 信号,每个音频声道被进行反向滤波以把信号转回到时域。

[0250] 接收音频帧和拆包头信息

[0251] 编码数据流是在编码器中打包(或成帧)的,它每帧中除了包括真正的音频码字本身还包含了附加数据,用于解码器同步、误差检测及纠正、音频编码状态标记值和编码辅助信息。拆包器 40 检测到 SYNC 字并取出帧尺寸 FSIZE。编码的比特流由连续的音频帧组成,每一帧都以 32-位 (0x7ffe8001) 同步字 (SYNC) 开始。音频帧的实际尺寸 FSIZE 从同步字之后的字节中取出。这样便允许程序员设定一个‘帧结束’定时器以减少软件运算代价。接着取出 Nblks 使解码器能算出音频窗口尺寸 (32(Nblks+1))。这告知解码器取出什么样的辅助信息和产生多少重构采样。

[0252] 帧头信息字节组 (sync, ftype, surp, nblks, fsize, amode, sfreq, rate, mixt, dynf, dynct, time, auxcnt, lff, hflag) 一旦被收到,就可以用里德所罗门校验字节 HCRC 来验证最前 12 个字节的有效性。这些程序可在 14 个字节中发生 1 个错误字节时进行纠正,或在发生 2 个错误字节时进行警示。在完成了错误校验之后,这些头信息被用来更新解码器标记值。

[0253] 位于 HCRC 之后直至可选信息的部分是头信息参数 (filts, vernum, chist, pcmr, unspec),可被取出并用来更新解码器标记值。由于这些信息不会逐帧改变,所以可以用多数表决方案来补偿其比特误差。可选的头数据 (times, mcoeff, dcoeff, auxd, ocrc) 则可根据 mixct, dynf, time 和 auxcnt 等头参数来取出。可选数据可以使用可选的里德所罗门

校验字节 OCRC 来验证。

[0254] 音频编码帧中的头参数 (subfs, subs, chs, vqsub, joinx, thuff, shuff, bhuff, sel5, sel7, sel9, sel13, sel17, sel25, sel33, sel65, sel129, ahcrc) 每帧传输一次。它们可以使用音频里德所罗门校验字节 AHCRC 来验证。大多数头信息相对于每个音频声道重复, 音频声道数由 CHS 定义。

[0255] 拆包子帧编码辅助信息

[0256] 音频编码帧被分成多个子帧 (SUBFS)。每个子帧包含所有正确解出音频必需的辅助信息 (pmode, pvq, tmode, scales, abits, hfreq) 而无须参考任何其它子帧。每个连续子帧的解码首先通过拆包其辅助信息。

[0257] 对于所有音频声道的每一个有效子带, 一个 1- 比特的预测模式 (PMODE) 标记值将被传输。PMODE 标记值对当前子帧有效。PMODE = 0 意味着该子带的音频帧中不包含预测器系数。在这种情况下, 此频带的预测器系数在该子帧的时段内被置零。PMODE = 1 意味着辅助信息内包含了该子带的预测器系数。在这种情况下, 这些预测器系数被取出并在该子帧的时段内被装入用于其预测器中。

[0258] 对于 pmode 数据列中的每个 PMODE = 1, 数据列 PVQ 中含有其相应的预测系数 VQ 地址索引。这些索引值为固定的、无符号的 12- 比特整数, 通过把 12- 比特整数映射至矢量表 266, 可以从查寻表中取出 4 个预测系数。

[0259] 比特分配索引值 (ABIT) 表示在逆量化器中的层数, 逆量化器把子带音频代码转换成绝对值。每个音频声道中的 ABITs 其拆包格式各不相同, 决定于它的 BHUFF 索引和具体的 VABIT 代码 256。

[0260] 瞬态模式辅助信息 (TMODE) 238 被用来表示每个子带中瞬变在子帧中的位置。每个子帧被分成 1-4 个子子帧。从子带采样数量来说, 每个子子帧由 8 个采样构成。最大子帧尺寸是 32 个子带采样。如果瞬变发生在第一子子帧内, 则 tmode = 0。当 tmode = 1 时表示瞬变出现在第二子子帧内, 以此类推。为了控制瞬变失真如预回波等, TMODE 大于零的子帧子带将传输两个比例因数。从音频头参数中取出的 THUFF 索引值决定了用于解码 TMODEs 的方法。当 THUFF = 3 时, TMODEs 作为无符号 2- 比特整数拆包。

[0261] 比例因数索引值的传输使每个子帧内的子带音频码可以进行适当比例运算。如果 TMODE 等于零, 则传输一个比例因数。如果任何子带的 TMODE 大于零, 那么同时传输两个比例因数。从音频头参数中取出的 SHUFF 索引值 240 决定了每个不同音频声道用于解码 SCALES 的方法。VDRMS_{ql} 索引值确定了 RMS 比例因数的值。

[0262] 在某些模式下, SCALES 索引值的拆包需要从五个 129- 层带符号的霍夫曼逆量化器中选一进行。然而, 其产生的逆量化索引值仍为差分编码形式, 需按下列方法转换成绝对值:

[0263] $ABS_SCALE(n+1) = SCALES(n) - SCALES(n+1)$

[0264] 其中 n 是在音频声道中从第一子带开始数的第 n 个差分比例因数。

[0265] 在低比特码率音频编码模式下, 音频编码器利用矢量量化直接对高频子带音频采样进行有效编码。这些子带不使用差分编码, 与正常 ADPCM 过程有关的所有数据列必须保持置零状态。VQSUB 表示了用 VQ 编码的第一个子带, 在此之上的、直至 SUBS 的所有子带都用这种方式进行编码。

[0266] 高频索引值 (HFREQ) 被拆包 248 为固定的、10- 比特无符号整数。通过应用合适的索引值,从 Q4 分数式二进制 LUT 中可取出每个子带子帧所需的 32 个采样。这一过程对每个启用高频 VQ 模式的声道进行重复。

[0267] 低音效果声道的抽取因数总是 $X128$ 。LFE 中的 8- 比特效果采样的数量当 $PSC = 0$ 时可由 $SSC * 2$ 给出,或当 PSC 为非零值时由 $(SSC+1) * 2$ 给出。LFE 数据列之后还包含有一附加 7 比特比例因数 (无符号整数),用 7- 比特 LUT 可将其转换成 rms。

[0268] 拆包子子帧音频码数据列

[0269] 子带音频码的抽取过程是通过 ABIT 索引值来实现的,且在 $ABIT < 11$ 的情况下还需使用 SEL 索引值。音频代码采用可变长度的霍夫曼码或固定的线性代码进行格式。通常,低于或等于 10 的 ABIT 索引值意味着采用了霍夫曼可变长度编码,其码本由代码 $VQL(n)$ 258 选定,而大于 10 的 ABIT 总是代表采用了固定长度码。所有量化器都具有中点取值、均匀步长的特性。对于固定码 (Y2) 量化器而言,最低的负数量化层被放弃不用。音频码被打包成子子帧,每个子子帧代表了最多为 8 个子带采样值,且在当前子帧中最多可有 4 个子子帧。

[0270] 如果采样频率标记值 (SFREQ) 表示的采样频率高于 48kHz,那么音频帧中将存在超音频数据阵列。该数据列中的头两个字节将表示超音频的字节尺寸。此外,解码器硬件的采样频率应根据具体高频采样频率值设定成 $SFREQ/2$ 或 $SFREQ/4$ 进行工作。

[0271] 拆包同步校验

[0272] 在每个子帧结束时,应对数据拆包同步校验字 $DSYNC = 0xffff$ 进行检测以便验证拆包的完好性。如果头信息参数、辅助信息或音频数据列因比特错误而遭到破坏,在辅助信息和音频码字中使用可变长度代码字节 (即低比特码率音频的情形) 可能导致拆包数据错位。如果拆包指针没有指向 $DSYNC$ 的始端,那么可以认为前一个子帧的音频是不可靠的。

[0273] 一旦所有的辅助信息和音频数据都进行了拆包,解码器一次一个子帧地重构多通道音频信号。图 20 示出了用于单个声道中单个子带的基带解码器部分。

[0274] 重构 RMS 比例因数

[0275] 解码器对 ADPCM、VQ 和 JFC 算法重构 RMS 比例因数 (SCALES)。具体来说,VTMODE 和 THUFF 索引值被逆映射以识别当前子帧的瞬态模式 (TMODE)。此后,SHUFF 索引值、 $VDRMS_{QL}$ 代码和 TMODE 被逆映射以便重构差分 RMS 编码。差分 RMS 编码被进行逆差分编码 242 以便选择 RMS 码,RMS 码再被逆量化 244 以形成 RMS 比例因数。

[0276] 逆量化高频矢量

[0277] 解码器对高频矢量进行逆量化以重构子带音频信号。具体是,由起始 VQ 子带 ($VQSUBS$) 所指示而取出的、有符号的 8- 位分数式 (Q4) 二进制数,即高频采样 (HFREQ),被映射至逆 VQ 查值表 248。选中的表格值被逆量化 250,并通过 RMS 比例因数进行比例运算 252。

[0278] 逆量化音频码

[0279] 在进入 ADPCM 循环运算之前,音频码被进行逆量化和比例运算以形成重构子带差分采样。逆量化的实现首先通过逆映射 VABIT 和 BHUFF 索引值,以指定用于确定步长和量化层数量的 ABIT 索引值,同时逆映射用于产生量化器层代码 $QL(n)$ 的 SEL 索引值和 $VQL(n)$ 音频码。然后,代码字 $QL(n)$ 映射到由 ABIT 和 SEL 索引值指定的逆量化器查找表 260。虽然代码是按 ABIT 排列顺序,但是每个不同的音频声道都有其不同的 SEL 值。查找过程将产

生带符号的量化器层数,它们可通过乘以量化器步长被转换成单位 rms。这些单位 rms 值然后乘以指定的 RMS 比例因数 (SCALES) 262 可以被转换成最后差分采样值。

[0280] 1. $QL[n] = 1/Q[\text{码}[n]]$, 其中 $1/Q$ 是逆量化器查找表

[0281] 2. $Y[n] = QL[n] * \text{步长}[\text{abits}]$

[0282] 3. $Rd[n] = Y[n] * \text{比例因数}$, 其中 Rd = 重构的差分采样

[0283] 逆 ADPCM

[0284] ADPCM 解码过程按照下列方式对每个子带差分采样执行:

[0285] 1. 从逆 VQ 查值表 268 中装入预测系数。

[0286] 2. 将当前的预测器系数与保留在预测器历史数据列 268 中的前 4 个重构子带采样进行卷积运算, 产生预测采样。

[0287] $p[n] = \text{sum}(\text{Coeff}[i] * R[n-i])$, n = 当前采样周期, $i = 1, 4$

[0288] 3. 将预测采样与重构的差分采样相加产生重构的子带采样 270。

[0289] $R[n] = Rd[n] + p[n]$

[0290] 4. 更新预测器的历史, 即把当前的重构子带采样值复制到历史数据列的顶部。

[0291] $R[n-i] = R[n-i+1]$, $i = 4, 1$

[0292] 在 $PMODE = 0$ 的情况下, 预测器系数将为零, 预测采样也为零, 而且重构子带采样等于差分子带采样。虽然在这种情况下不需要进行预测计算, 重要的是仍需要保持预测器历史的更新, 以便 $PMODE$ 在将来的子帧中又重新启用。此外, 如果 $HFLAG$ 在当前音频帧中有效, 则应在解码该帧中第一个子子帧之前清除预测器历史。此点以后的历史则应该照常更新。

[0293] 对于高频 VQ 子带或非编码 (即位于 $SUBS$ 限值以上) 子带来说, 预测器历史应保持清零状态, 直到其子带预测器被启用为止。

[0294] ADPCM、VQ 和 JFC 解码的选择控制

[0295] 第一个“开关”控制了 ADPCM 或 VQ 输出的选择。VQSUBS 索引值识别出 VQ 编码的起始子带。因此, 如果当前子带低于 VQSUBS, 开关将选择 ADPCM 输出。否则它将选择 VQ 输出。第二个“开关”278 控制了直接声道输出或 JFC 编码输出的选择。JOINX 索引值确定了哪些声道被联合及重构信号是在哪一个声道中产生的。重构的 JFC 信号形成了其它声道中 JFC 输入的强度源。因此, 如果当前子带是 JFC 的一部分而且不是指定声道, 则开关将选择 JFC 输出。一般情况下开关选择声道输出。

[0296] 下混音矩阵

[0297] 数据流的音频编码模式由 $AMODE$ 指出。解码的音频声道然后可被重新定向以与解码器硬件 280 上的实际输出声道安排相匹配。

[0298] 动态范围控制数据

[0299] 在编码阶段 282, 动态范围系数 $DCOEFF$ 可被选择性地嵌入音频帧中。这个特性的目的是便于在解码器的输出中实现音频动态范围的压缩。动态范围的压缩对于某些收听环境尤为重要, 这些环境的高环境噪声级使低能量信号不能被察觉, 除非冒着在高音量部分损坏扬声器的危险。20- 比特 PCM 音频录音技术的日益广泛应用使该问题进一步复杂化, 此类录音具有高达 110dB 的动态范围。

[0300] 根据帧的窗口尺寸 (NBLKS), 无论哪种编码模式, 每个声道可传输一、二或四个系

数 (DYNF)。如果传输单个系数,则它可用于整个帧。如果传输两个系数,则第一个系数用于帧的前半部分,而第二个系数用于帧的后半部分。四个系数则分布在每个帧的四等份上。更高的时间分辨率可在本地对传输值进行内插运算达到。

[0301] 每个系数是 8- 位带符号的分数式 Q2 二进制数,而且代表了如表 (53) 所示出的对数增益值,该表给出范围为 $\pm 31.75\text{dB}$,步长为 0.25dB 。这些系数按照声道数进行排序。动态范围压缩通过将解码音频采样乘以这些线性系数达到。

[0302] 解码器通过适当调节这些系数值可以改变压缩的程度,或可以忽略这些系数而完全中止动态范围压缩。

[0303] 32- 频带内插滤波器组

[0304] 32- 频带内插滤波器组 44 将每个声道的 32 个子带转换成单个 PCM 时域信号。当 FILTS = 0 时使用非完全重构系数 (512- 节 FIR 滤波器)。当 FILTS = 1 时使用完全重构系数。通常余弦调制系数可预先计算并存储在 ROM (只读存储器) 中。内插程序可扩展用于重构大尺寸数据块以降低循环程序的代价。然而,在终止帧的情况下,需要的最小分辨率的是 32 个 PCM 采样。内插算法如下:建立余弦调制系数,将 32 个新子带采样读入数据列 XIN,乘以余弦调制系数和建立临时数据列 SUM 和 DIFF,存储历史值,乘以滤波器系数,建立 32 个 PCM 输出采样,更新工作数据列,和输出 32 个新 PCM 采样。

[0305] 根据所使用的比特码率和编码方案,比特流既可以指定非完全或完全重构内插滤波器组系数 (FILTS)。由于编码器抽取滤波器组是用 40- 比特浮点精度来计算的,所以编码器达到最大理论重构精度的能力取决于源 PCM 字长和 DSP 核心用于计算卷积的精度以及操作中比例运算的方式。

[0306] 低频率效果 PCM 内插

[0307] 与低频效果声道相关的音频数据独立于主音频声道。该声道是用 8- 比特 APCM 过程对 X128 抽取的 (120Hz 带宽) 20- 比特 PCM 输入进行编码。抽取的效果音频在时间上需与主音频声道的当前子帧音频对齐。因此,由于所有 32- 频带内插滤波器组的延迟是 256 个采样 (512- 节),所以必须注意确保内插的低频效果声道在输出前也与其它音频声道对齐。如果效果内插 FIR 也是 512 节,则不需要进行补偿。

[0308] LFT 算法使用 512 节的 128X 内插 FIR 进行如下操作:把 7- 比特比例因数映射成 rms,乘以 7- 比特量化器的步长,由归一化值产生子采样值,以及利用低通滤波器进行内插 128,例如为每个子采样设置的那样。

[0309] 硬件实现

[0310] 图 21 和 22 描述了一个 6 声 (通) 道编码器和解码器硬件实现的基本功能结构,它可在 32、44.1 和 48kHz 等采样频率进行操作。参照图 22,八个模拟器件公司的 ADSP21020 40- 比特浮点数字信号处理器 (DSP) 芯片 296 被用来实现一个 6 声道数字音频编码器 298。六个 DSPs 中每个被用来对所有声道的其中之一进行编码,第七和第八个 DSP 则用来分别实现“全局比特分配与管理”和“数据流格式化及误差编码”功能。每个 ADSP21020 以 33MHz 的时钟驱动,并使用外接的 (48 比特 X32k) 程序读写存储器 (PRAM) 300, (40 比特 X32k) 数据读写存储器 (SRAM) 302 进行算法运算。在编码器情况下,还用了 (8 比特 X512k) 的 EPROM 304 来存储固定常数,例如可变长度的熵代码本。用于数据流格式化的 DSP 使用了里德所罗门 CRC (循环冗余校验) 芯片 306 以使解码器可以进行检错与纠错操作。编码器 DSPs 和全

局比特分配及管理之间的信息交流是通过双端口静态读写存储器

[0311] (RAM) 308 实现的。

[0312] 编码处理的流程如下。三个 AES/EBU 数字音频接收器的任意之一可输出一个 2-声道数字音频 PCM 数据流 310。这些双声道数据流中的第一声道分别导向 CH1、3 和 5 编码器 DSPs, 其第二声道分别导向 CH2、4 和 6。串行 PCM 字被转换成并行 (s/p) 以将 PCM 采样读入 DSPs。如前所述, 每个编码器累积一帧 PCM 采样然后对此帧数据进行编码。每个声道中有关估计差分信号 ($ed(n)$) 和子带采样 ($x(n)$) 的信息通过双端口 RAM 传输到全局比特分配和管理 DSP 中。然后用同样的方式读回每个编码器的比特分配策略。编码过程完成之后, 6 个声道的编码数据和辅助信息通过全局比特分配和管理 DSP 传至数据流格式化 DSP 中。在这一阶段, CRC 校验字节可被选择性地产生并加到编码数据中以便在解码器中提供误差保护。最后, 整个数据包 16 被组装输出。

[0313] 图 22 描述了一个 6 声道解码器的硬件实现。单个模拟器件公司的 ADSP21020 40-比特浮点数字信号处理器 (DSP) 芯片 324 被用来实现 6 声道数字音频解码器。此 ADSP21020 以 33MHz 的时钟驱动, 并使用外接的 (48 比特 X32k) 程序读写存储器 (PRAM) 326, (40 比特 X32k) 数据读写存储器 (SRAM) 328 运行解码算法。此外还用了附加的 (8 比特 X512k) EPROM 330 来存储诸如可变长度熵和预测系数矢量代码本等固定常数。

[0314] 解码过程的流程如下。压缩数据流 16 通过串行 / 并行转换器 (s/p) 332 被输入到 DSP。数据如前面所述被进行拆包和解码。每个通道的子带采样被重构成单个 PCM 数据流 22 并通过三个并行 / 串行转换器 (p/s) 335 输出到三个 AES/EBU 数字音频输出芯片 334 中。

[0315] 以上显示和描述了本发明的几个说明性实施例, 但是对于那些熟悉本领域的技术人员来说可以做出大量改变和替换的实施例。例如, 随着处理器速度的增加和存储器成本的降低, 采样频率、传输率和缓冲器尺寸很可能增加。这些变动的和替换的实施例可以被预想和实现, 但并没有脱离本发明如修正要求书中所定义构思和范围。

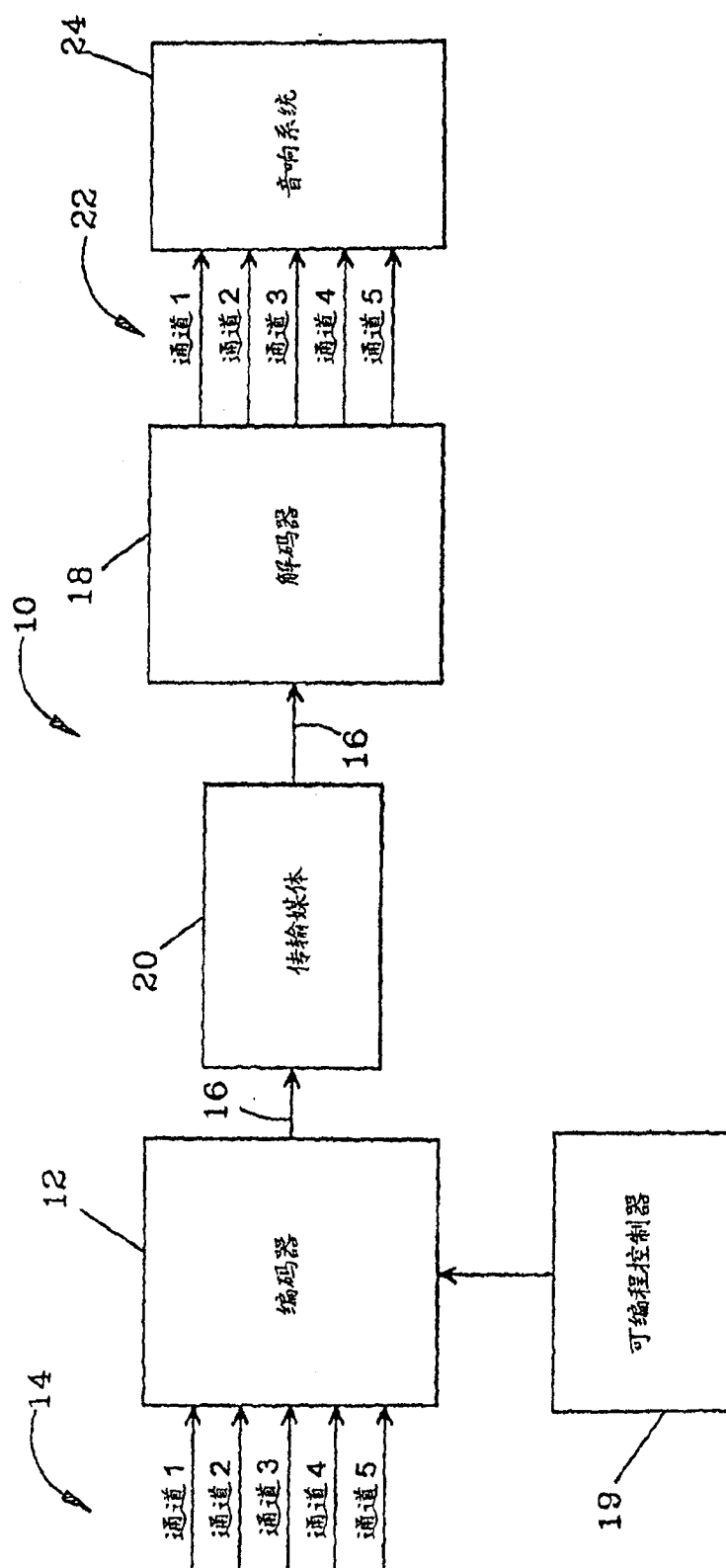


图 1

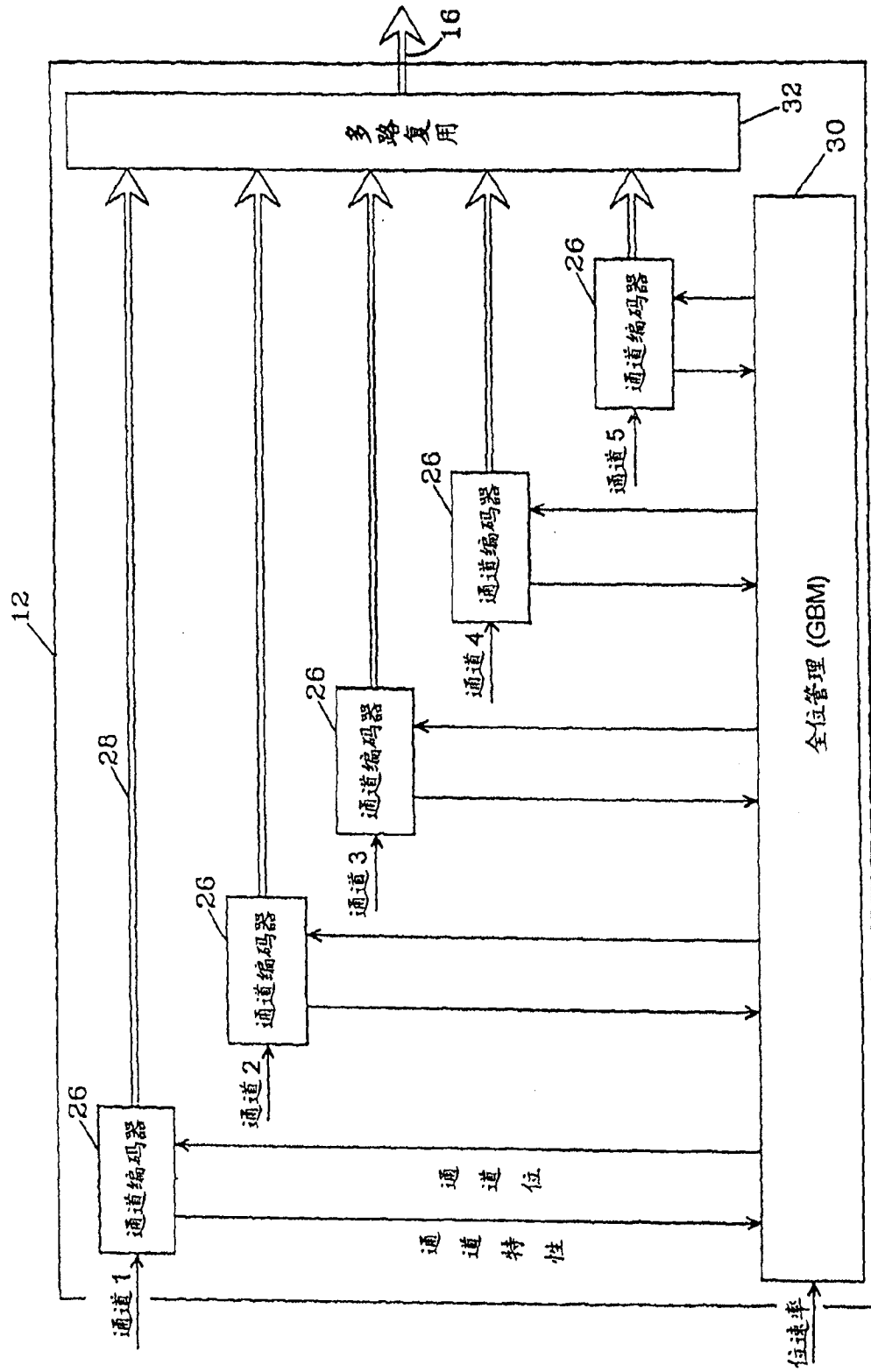


图 2

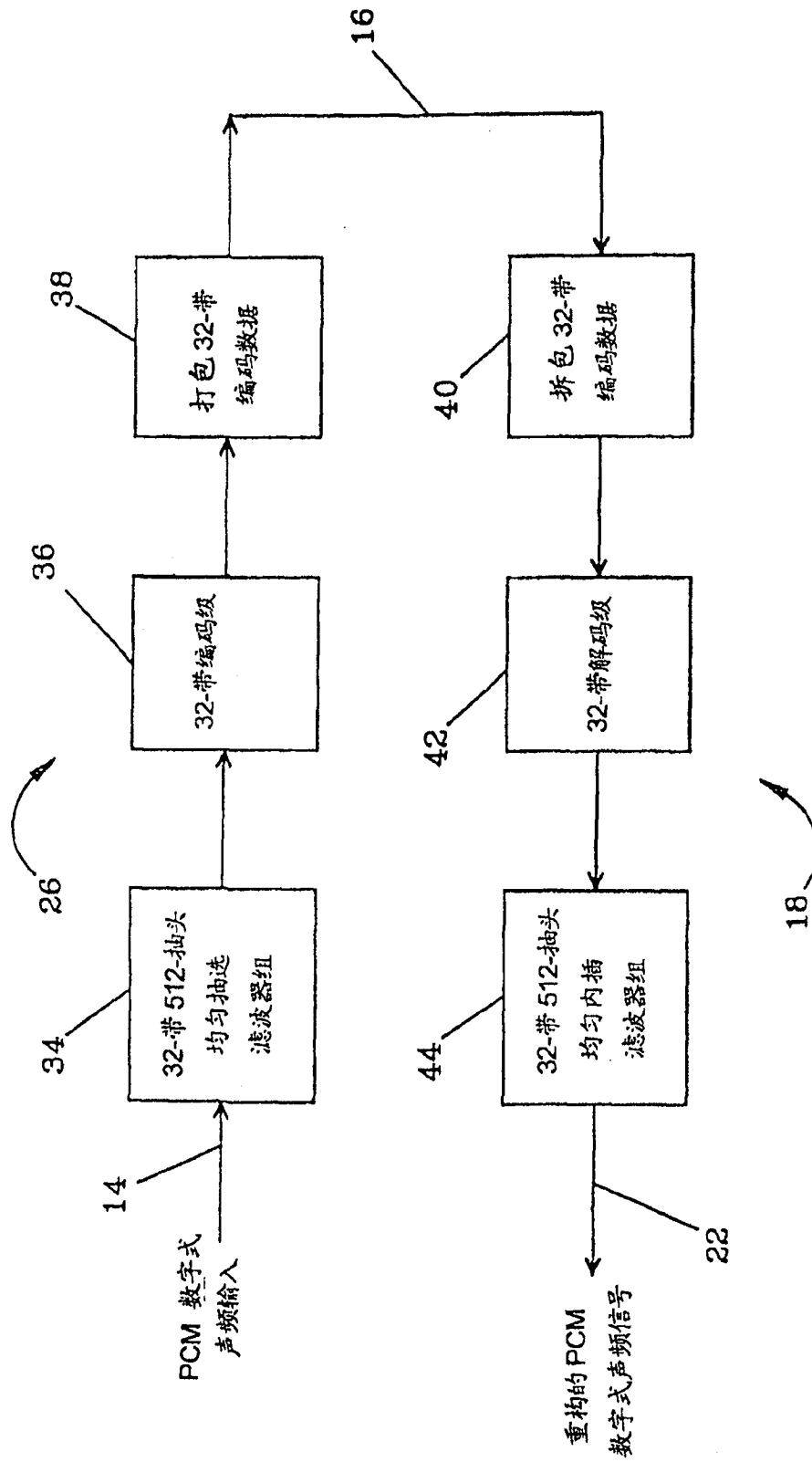


图 3

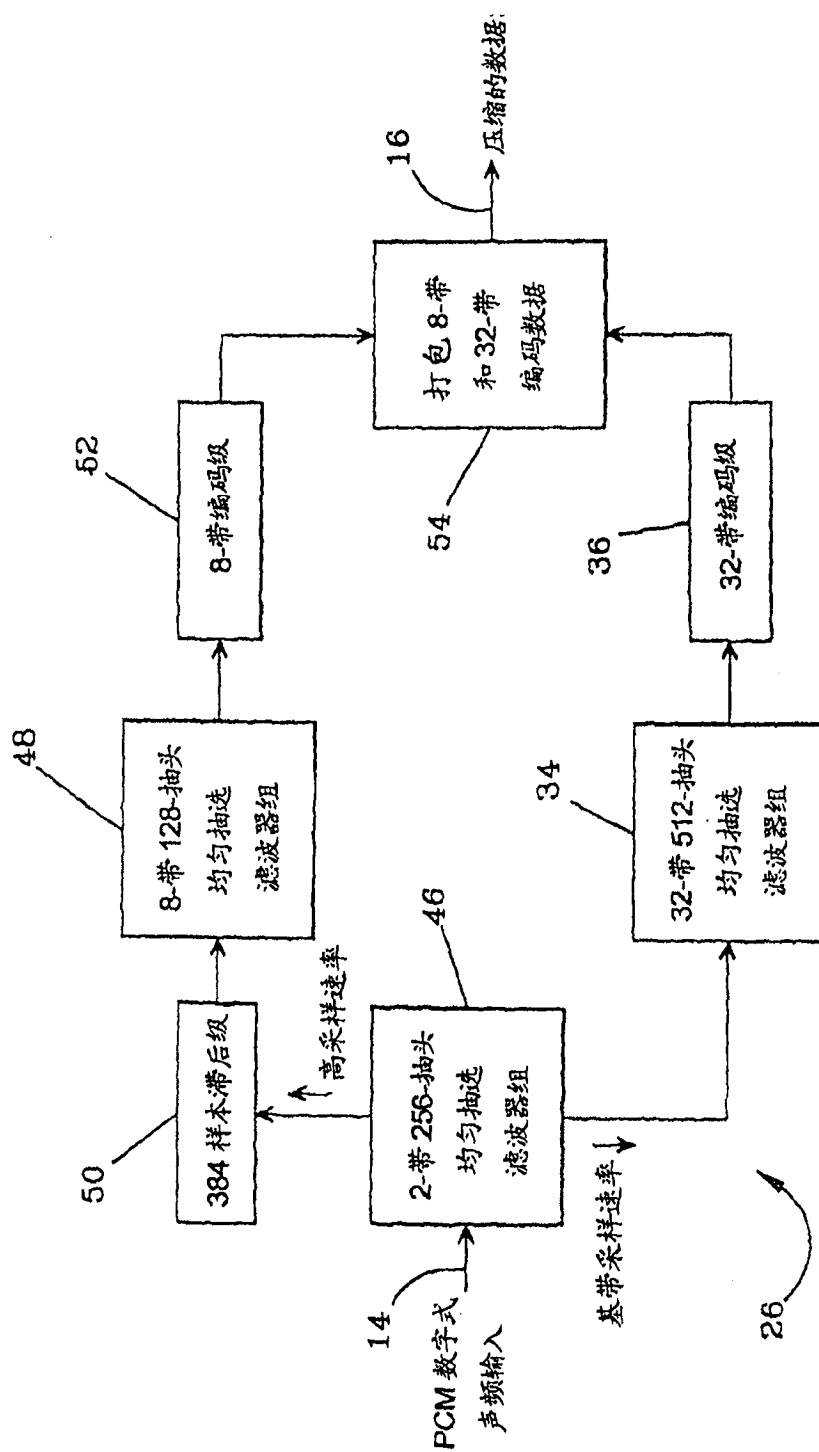


图 4a

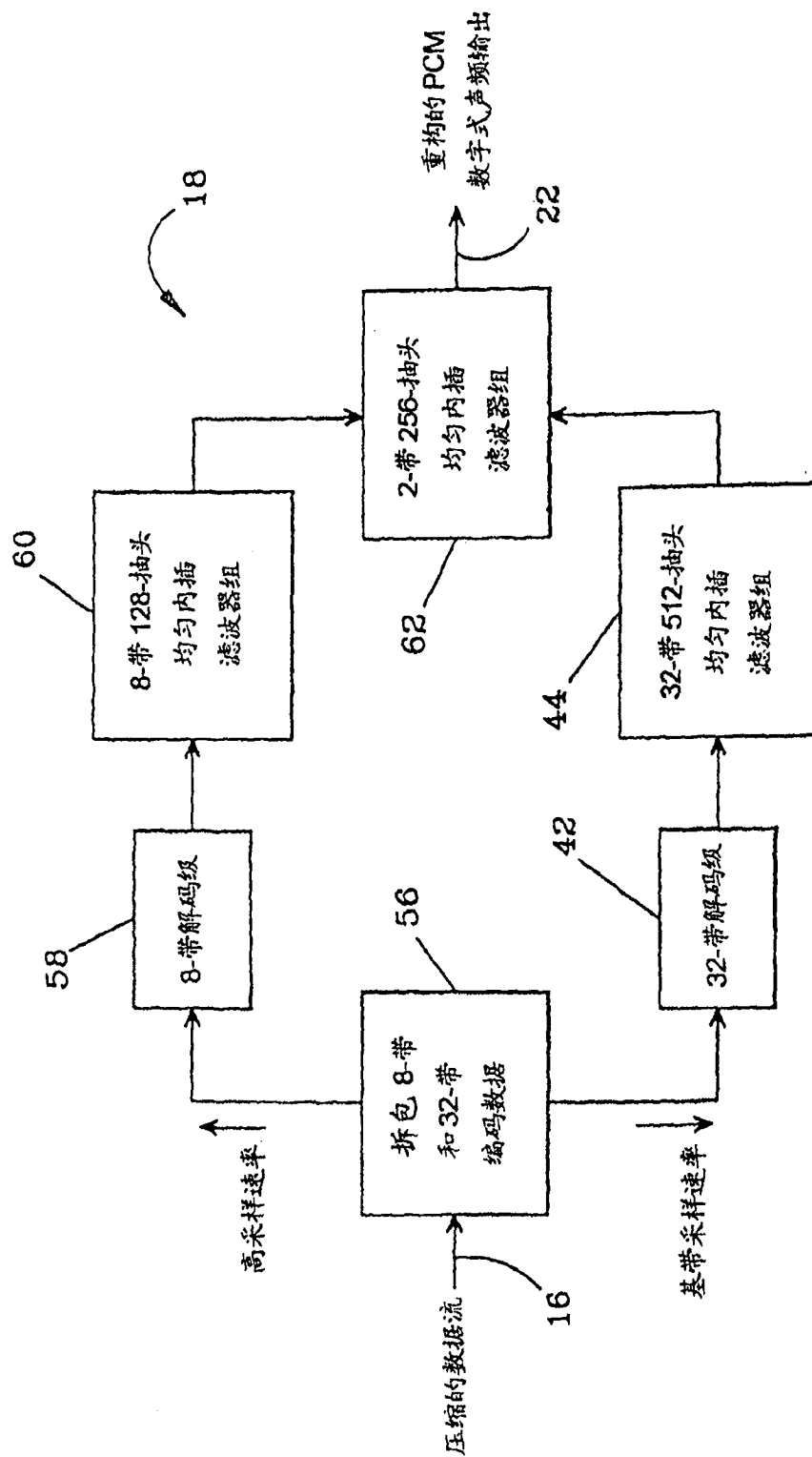


图 4b

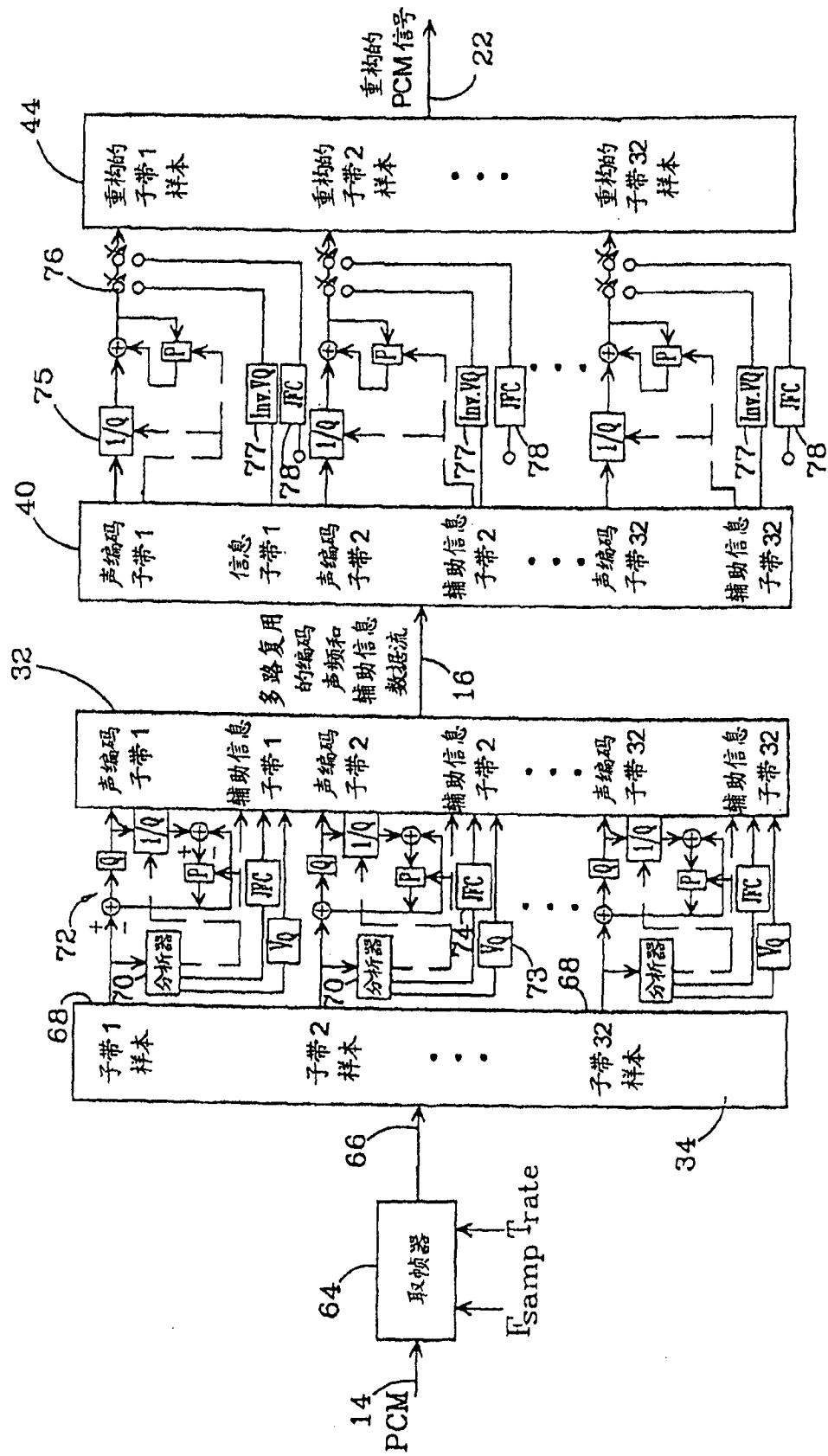


图 5

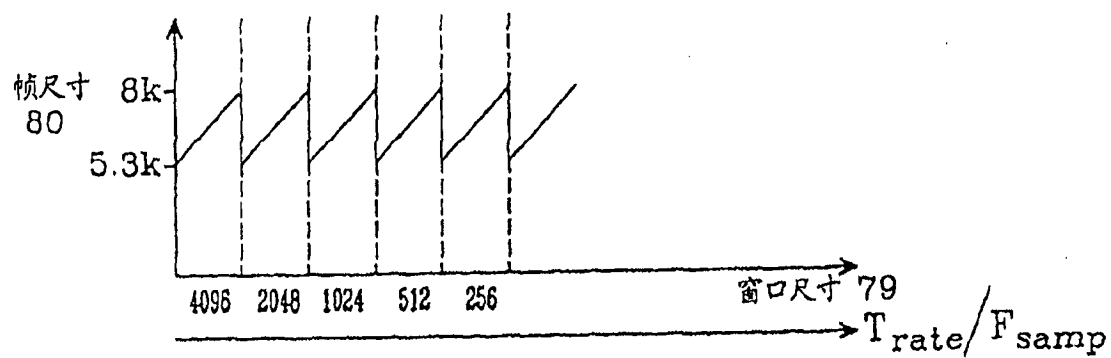


图 6

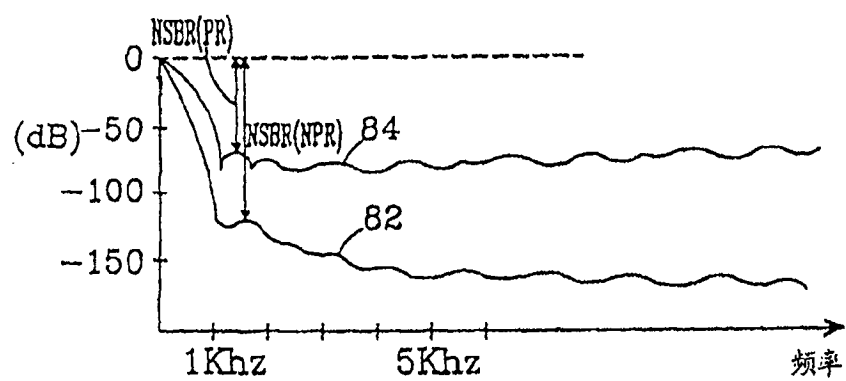


图 7

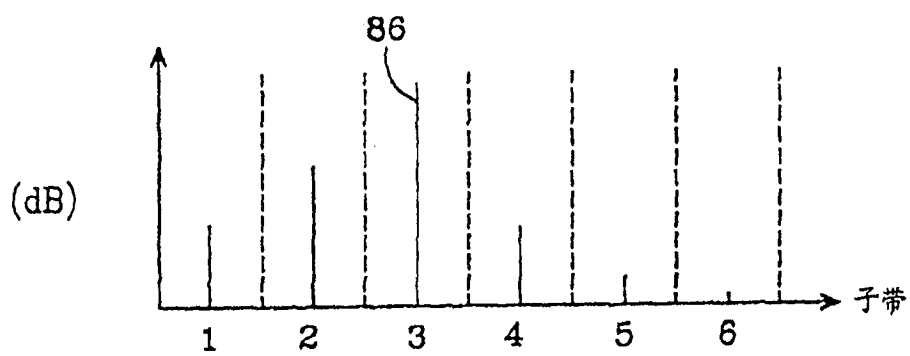


图 8

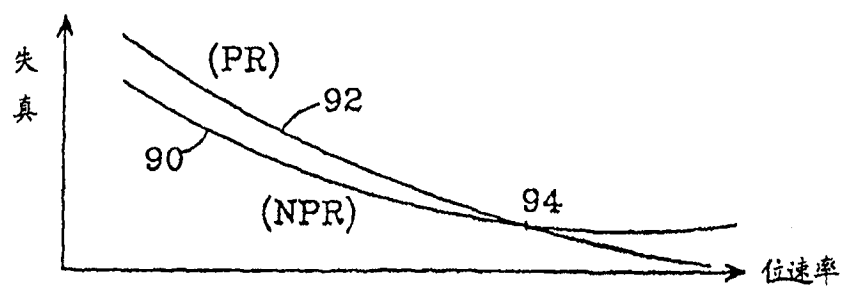


图 9

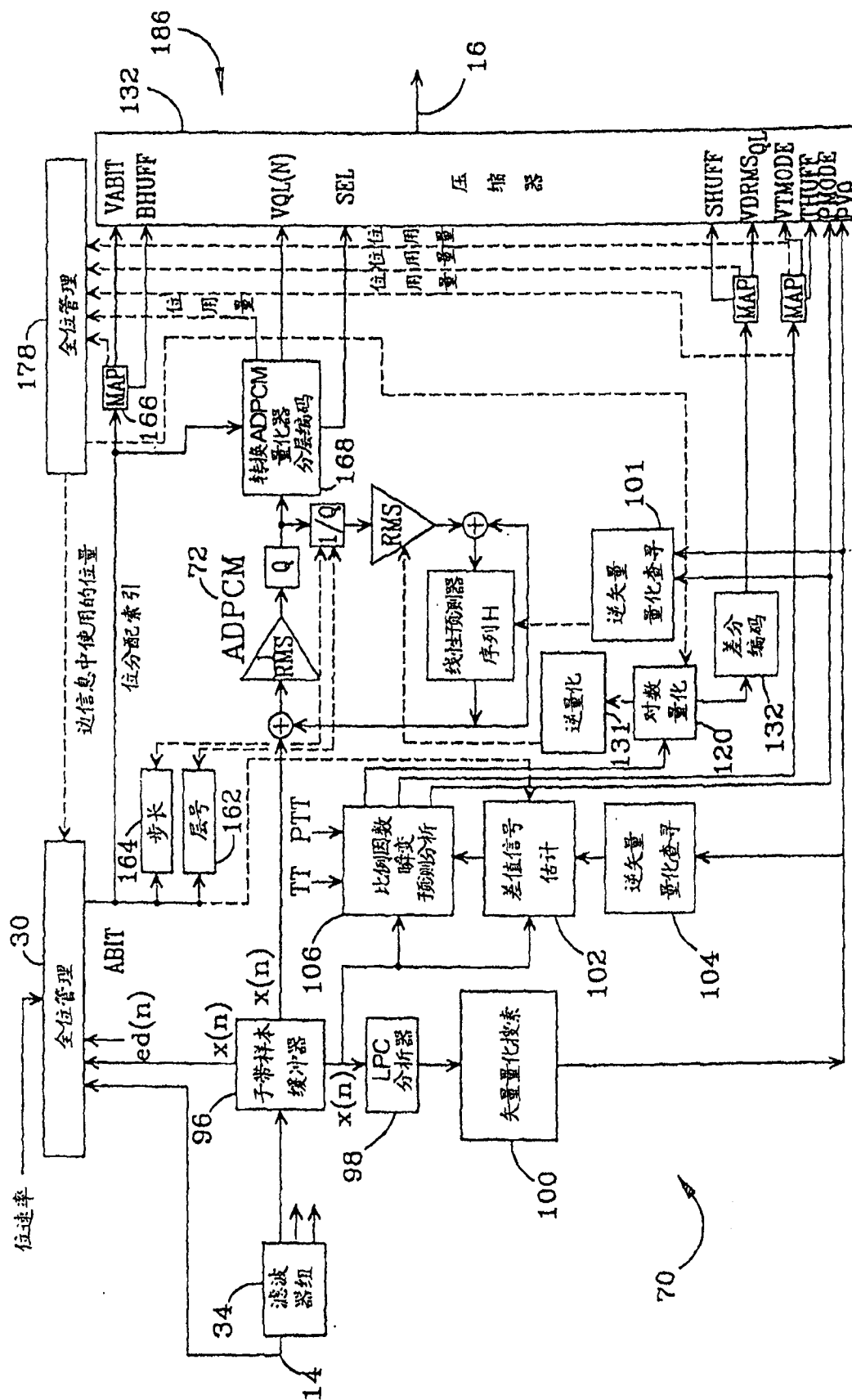


图 10

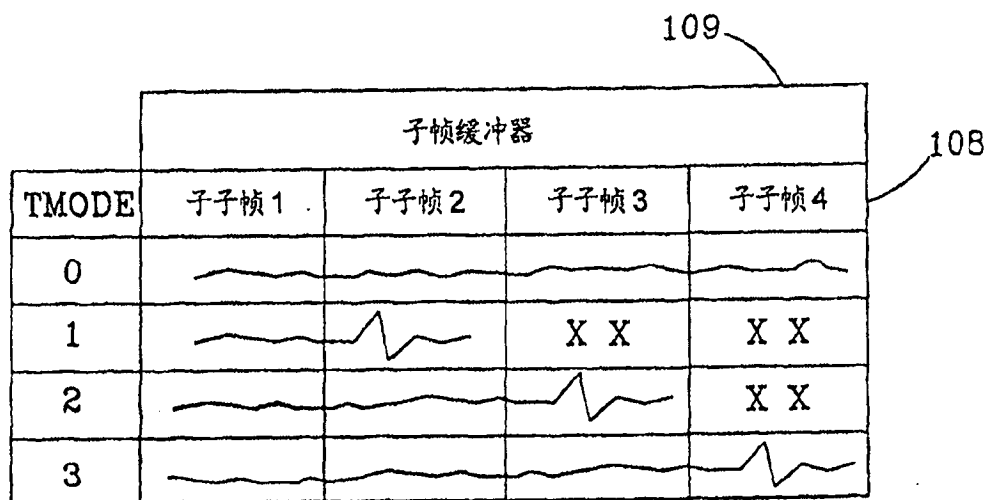


图 11A

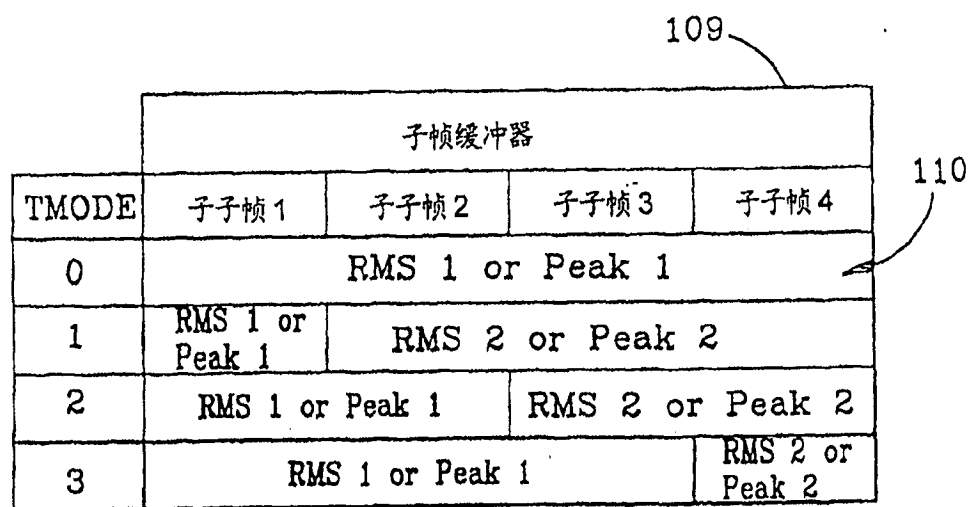


图 11B

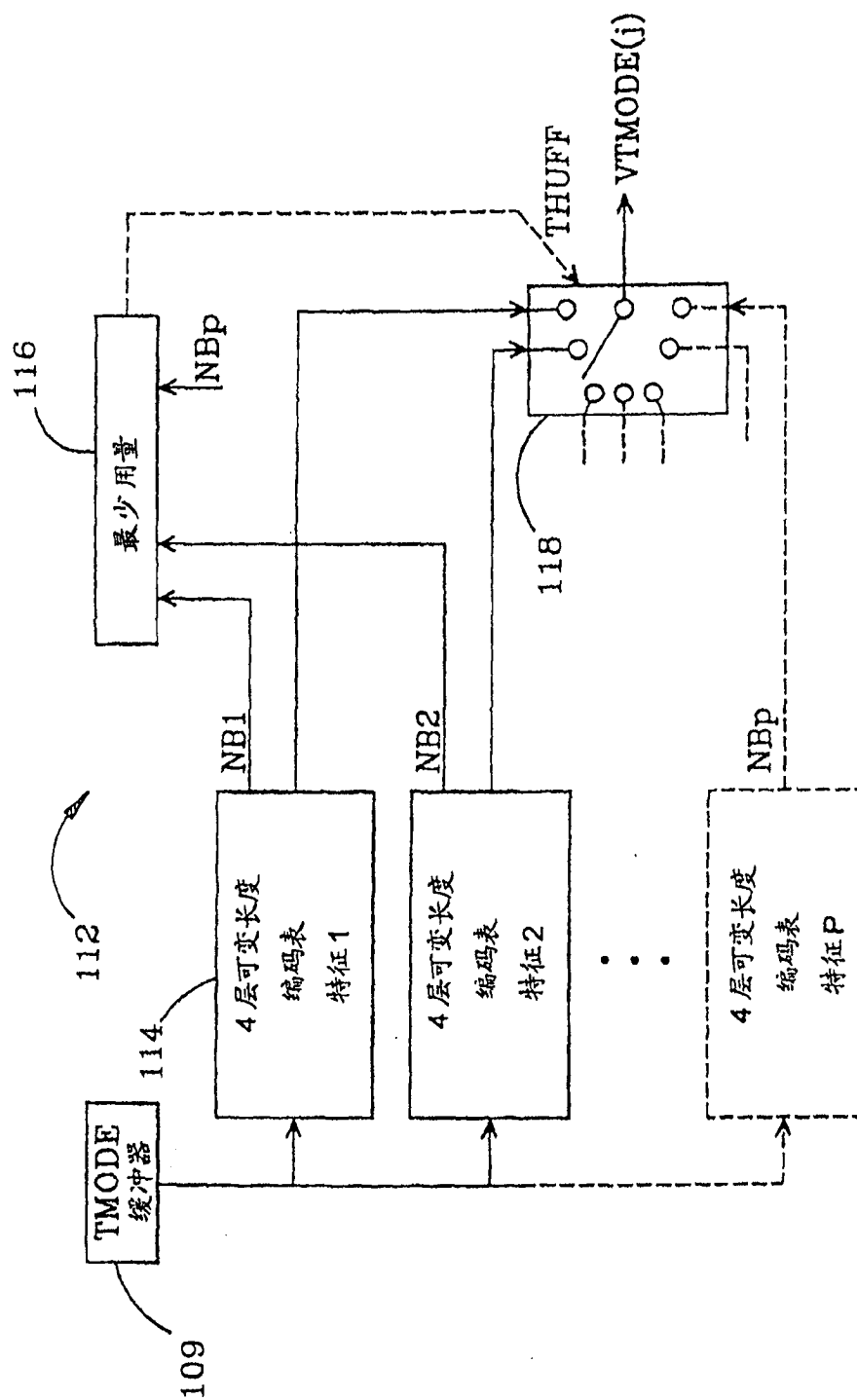


图 12

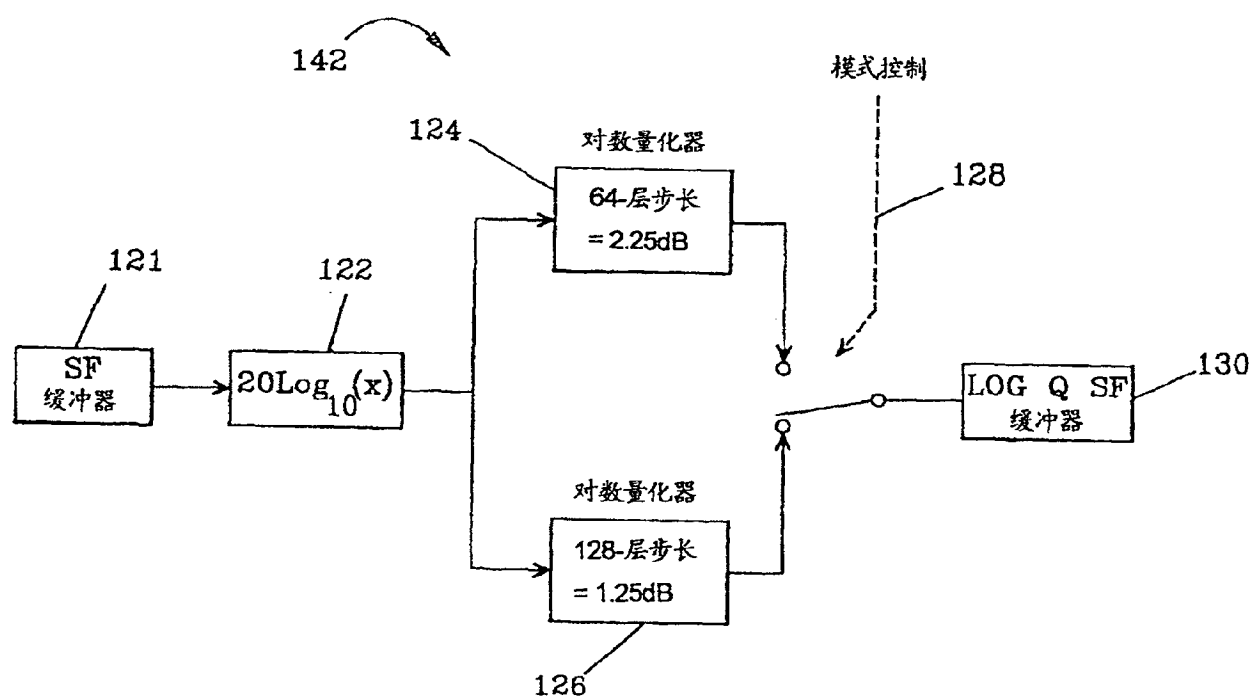


图 13

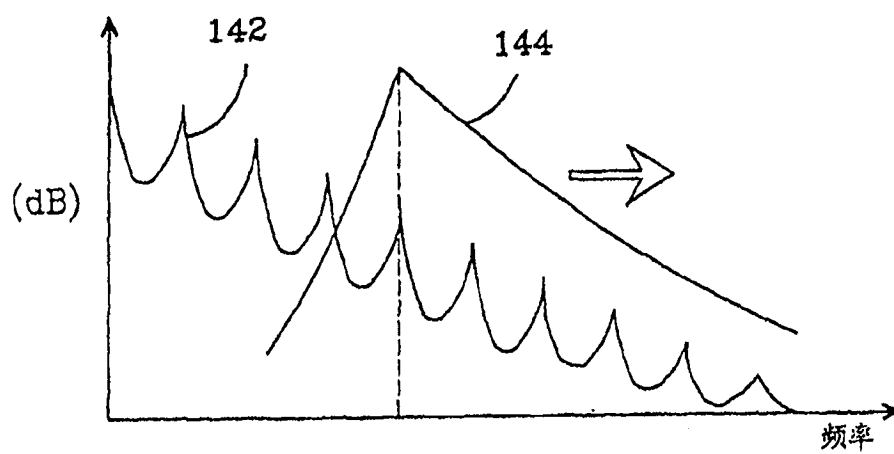


图 14

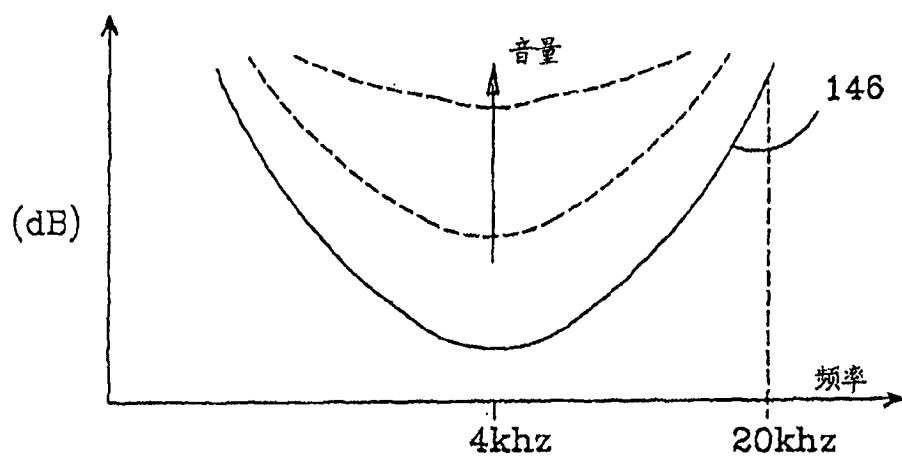


图 15

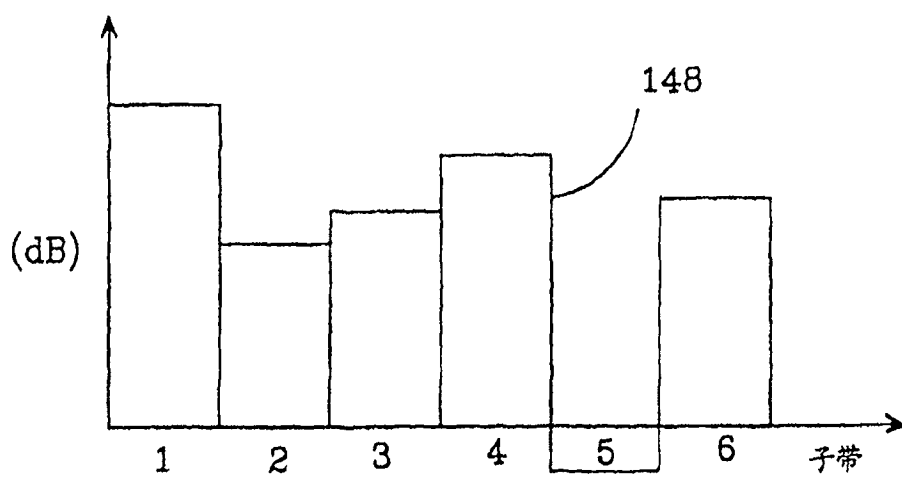


图 16

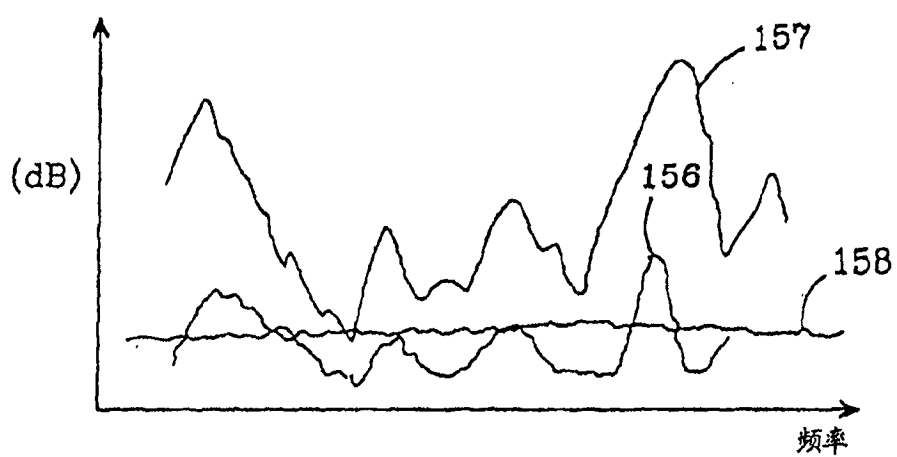


图 17

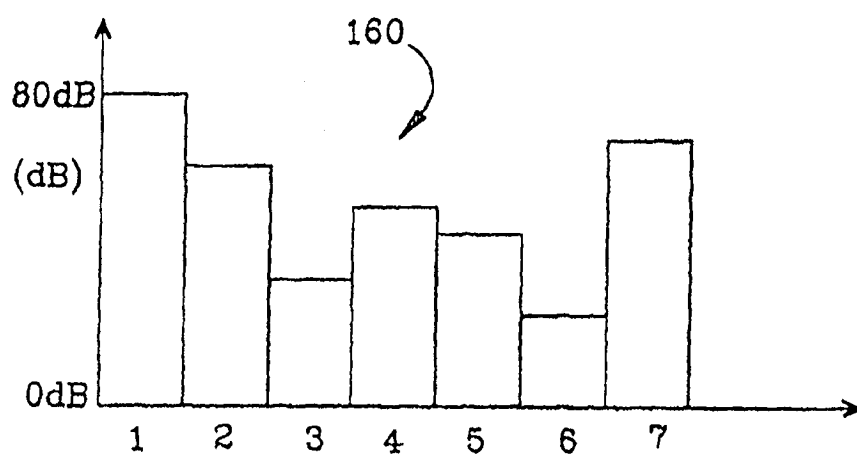


图 18A

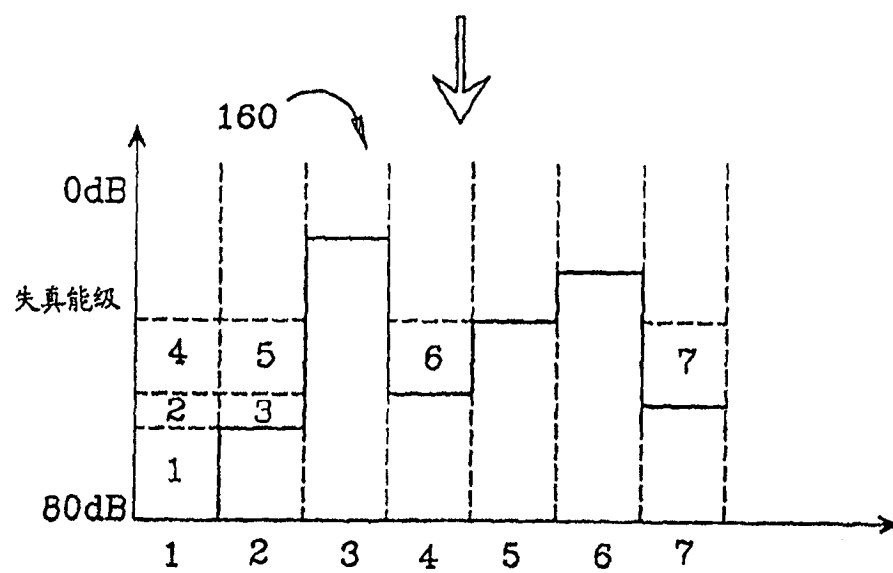


图 18B

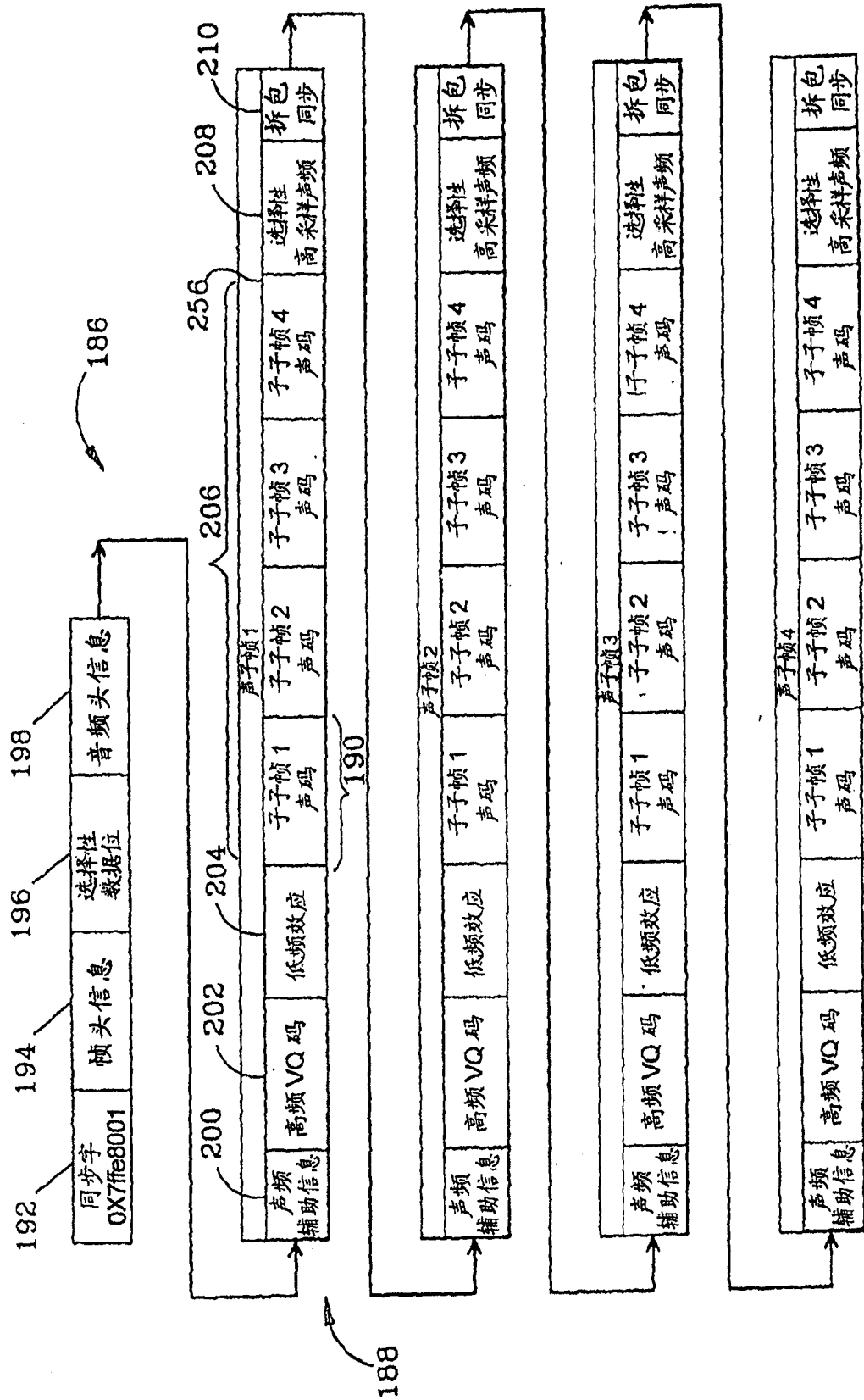


图 19

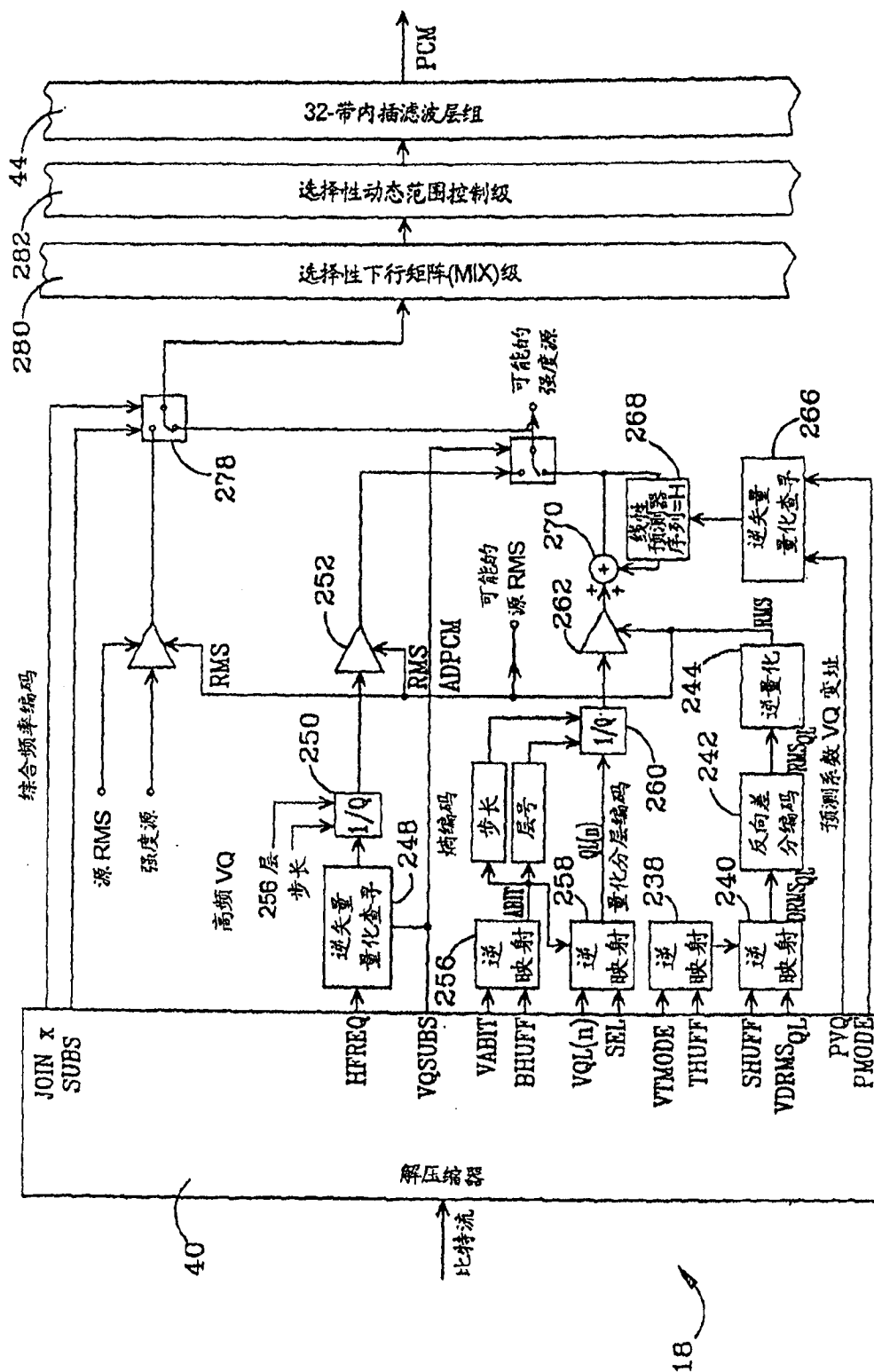


图 20

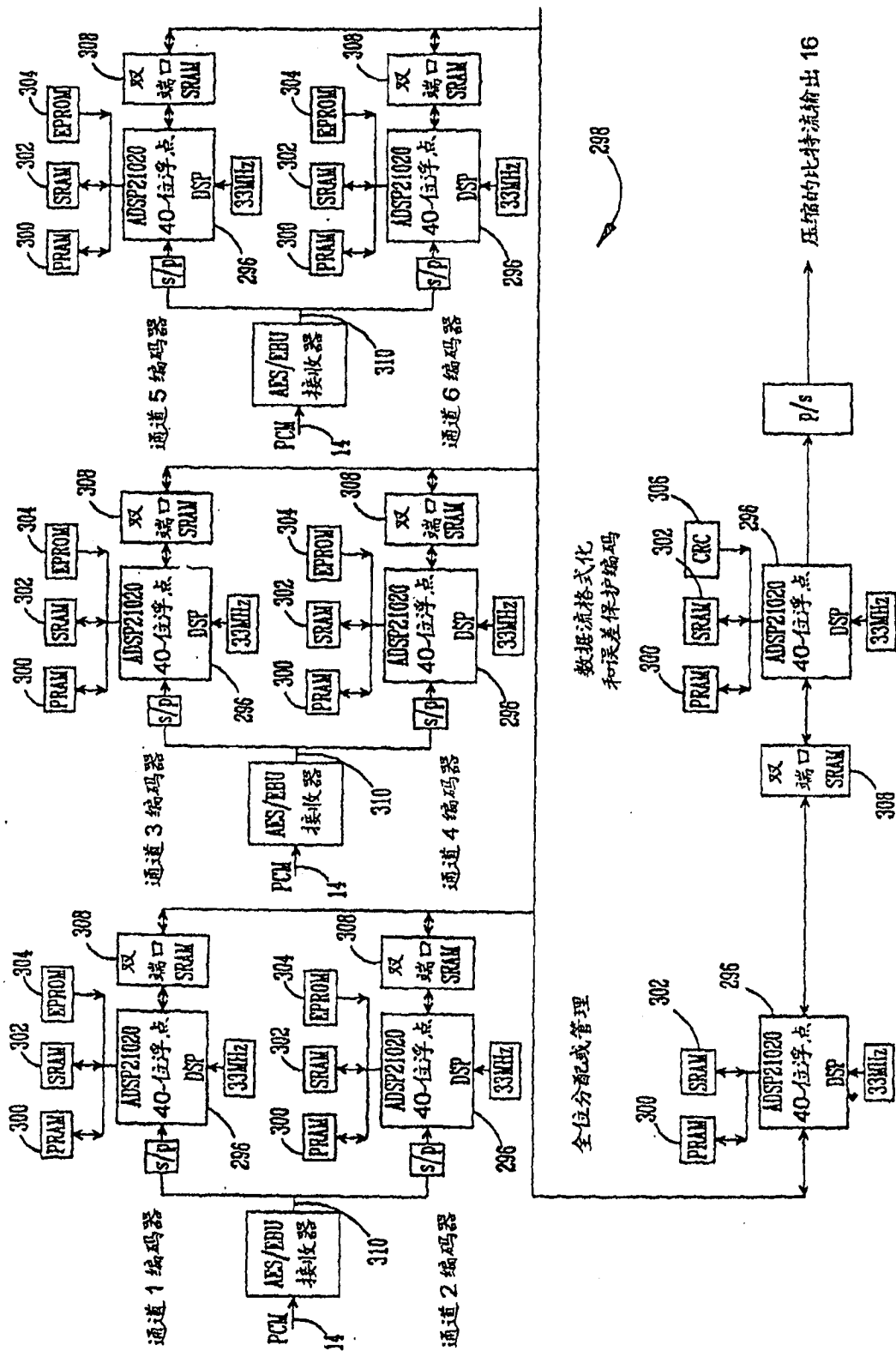


图 21

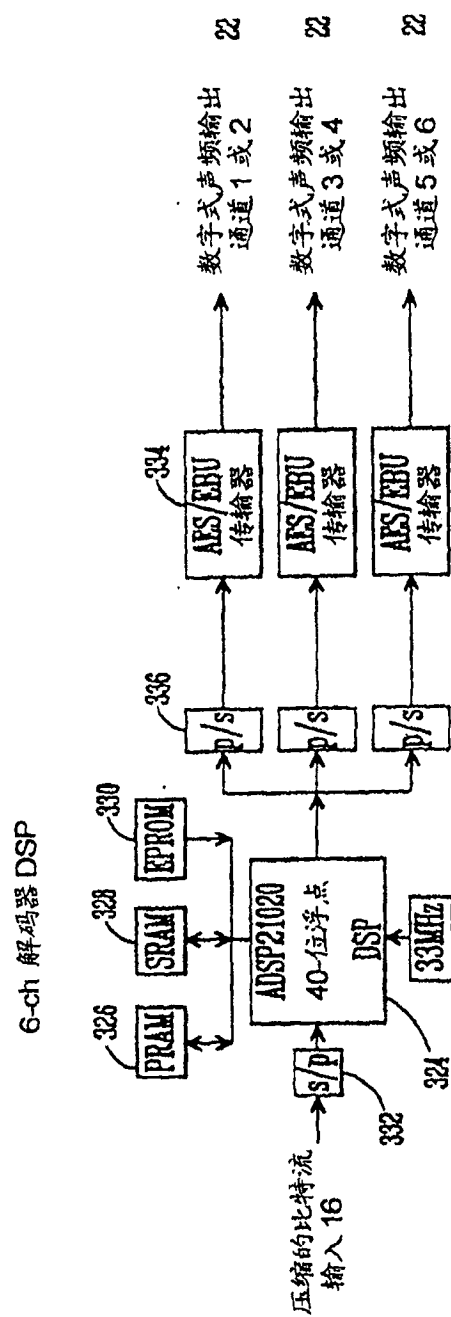


图 22