



(12)发明专利

(10)授权公告号 CN 105580075 B

(45)授权公告日 2020.02.07

(21)申请号 201480041566.7

(22)申请日 2014.07.15

(65)同一申请的已公布的文献号
申请公布号 CN 105580075 A

(43)申请公布日 2016.05.11

(30)优先权数据

13177350.9 2013.07.22 EP

13177353.3 2013.07.22 EP

13177348.3 2013.07.22 EP

13177346.7 2013.07.22 EP

13189368.7 2013.10.18 EP

(85)PCT国际申请进入国家阶段日
2016.01.22

(86)PCT国际申请的申请数据
PCT/EP2014/065116 2014.07.15

(87)PCT国际申请的公布数据
W02015/010952 EN 2015.01.29

(73)专利权人 弗劳恩霍夫应用研究促进协会
地址 德国慕尼黑

(72)发明人 克里斯蒂安·诺伊坎 萨沙·迪施
弗雷德里克·纳格尔
安德烈娅斯·尼德迈尔

康斯坦丁·施密特

巴拉吉·纳根德兰·托斯卡纳

(74)专利代理机构 北京集佳知识产权代理有限公司 11227

代理人 康建峰 杨华

(51)Int.Cl.

G10L 21/0388(2013.01)

G10L 19/02(2013.01)

(56)对比文件

CN 1813286 A, 2006.08.02,

CN 1647154 A, 2005.07.27,

CN 1918632 A, 2007.02.21,

CN 1918631 A, 2007.02.21,

CN 1114122 A, 1995.12.27,

CN 103165136 A, 2013.06.19,

CN 101622669 A, 2010.01.06,

SANJEEV MEHROTRA ET AL. "Hybrid low

bit rate audio coding using adaptive gain
shape vector quantization".《2008 IEEE
10th workshop on Multimedia signal
processing》.2008,

审查员 董小东

权利要求书4页 说明书39页 附图29页

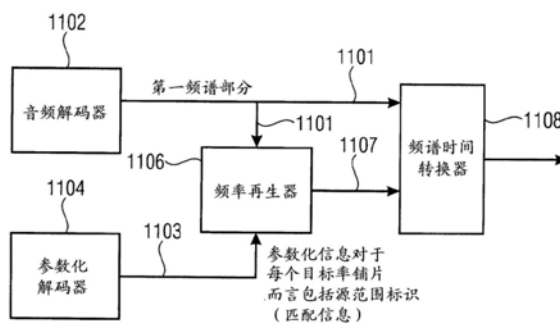
(54)发明名称

用自适应频谱铺片选择的音频信号解码和
编码设备及方法

(57)摘要

一种用于对编码信号进行解码的设备,包
括:音频解码器(1102),音频解码器(1102)用于
对第一组第一频谱部分的编码表示进行解码以
获得经解码的第一组第一频谱部分(1101);参数
化解码器(1104),参数化解码器(1104)用于对第
二组第二频谱部分的编码参数化表示进行解码
以获得参数化表示的解码表示(1103),其中,参
数化信息对于每个目标频率铺片而言包括源区
域标识作为匹配信息;以及频率再生器(1106),

频率再生器(1106)用于使用由匹配信息标识的
来自第一组第一频谱部分(1101)的源区域来再
生目标频率铺片。



1. 一种用于对编码信号进行解码的设备,包括:

音频解码器(1102),所述音频解码器(1102)用于对第一组第一频谱部分的编码表示进行解码以获得经解码的第一组第一频谱部分(1101);

参数化解码器(1104),所述参数化解码器(1104)用于对第二组第二频谱部分的编码参数化表示进行解码以获得所述参数化表示的解码表示(1103),其中,所述参数化表示的所述解码表示(1103)对于每个目标频率铺片而言包括源区域标识(1121)作为匹配信息;以及

频率再生器(1106),所述频率再生器(1106)用于使用由所述匹配信息标识的来自所述第一组第一频谱部分(1101)的源区域来再生目标频率铺片,

其中,所述频率再生器(1106)包括可控制的白化滤波器(1122,1123),

其中,所述参数化表示的所述解码表示(1103)包括白化信息(1123),

其中,所述频率再生器(1106)被配置用于:当根据所述匹配信息选择的源区域的所述白化信息指示所述源区域要被白化时,在执行频谱包络调整(1128)之前将所述白化滤波器(1122)应用于所述源区域,并且

其中,应用所述白化滤波器被执行以使得粗糙的包络信息被去除并且频谱精细结构被强调,或者其中,应用所述白化滤波器包括计算所述源区域的频谱包络估计以及将所述源区域的频谱除以由所述频谱包络估计指示的频谱包络。

2. 根据权利要求1所述的设备,

其中,所述音频解码器(1102)是频谱域音频解码器,并且

其中,所述设备还包括频谱时间转换器(1108),所述频谱时间转换器(1108)用于将所述第一频谱部分和经重构的第二频谱部分的频谱表示转换成时间表示。

3. 根据权利要求1所述的设备,

其中,所述白化信息对于铺片或一组铺片而言包括白化水平信息,所述白化水平信息指示当再生所述目标频率铺片时要被应用于源频率铺片的白化水平,并且

其中,所述频率再生器(1106)被配置用于:应用响应于所述白化信息从一组不同的白化滤波器中选择的白化滤波器。

4. 根据权利要求1所述的设备,

其中,所述频率再生器(1106)包括源区域修改器,

其中,所述参数化表示的所述解码表示(1103)除了包括所述源区域标识(1121)以外还包括符号信息,并且

其中,所述源区域修改器被配置用于:应用根据所述符号信息获得源区域频谱值的相移的操作。

5. 根据权利要求1所述的设备,

其中,所述频率再生器(1106)包括铺片调制器(1120),

其中,所述参数化表示的所述解码表示(1103)除了包括所述源区域标识(1121)以外还包括相关性滞后,并且

其中,所述铺片调制器被配置用于:根据与所述源区域标识(1121)相关联的所述相关性滞后来应用铺片调制。

6. 根据权利要求1所述的设备,

其中,所述频率再生器(1106)包括铺片调制器(1120),

其中,所述参数化表示的所述解码表示(1103)除了包括所述源区域标识(1121)以外还包括相关性滞后,并且

其中,所述铺片调制器(1120)被配置用于:当所述相关性滞后是奇数时,应用使用了交替的时间序列-1/1的铺片调制。

7.一种用于对音频信号进行编码的设备,包括:

时间频谱转换器(1130),所述时间频谱转换器(1130)用于将音频信号转换成频谱表示;

频谱分析器,所述频谱分析器用于分析所述频谱表示以确定要以第一频谱分辨率编码的第一组第一频谱部分和要以第二频谱分辨率编码的第二组第二频谱部分,其中,所述第二频谱分辨率低于所述第一频谱分辨率;

参数计算器(1134),所述参数计算器(1134)用于计算预定源区域与目标区域之间的相似度,源区域包括频谱部分,且目标区域包括第二频谱部分,

其中,所述参数计算器(1134)被配置用于:比较第一频谱部分和第二频谱部分的不同对的匹配结果以确定所选择的匹配对,并且提供标识所述匹配对的匹配信息;以及

核心编码器(1136),所述核心编码器(1136)用于对所述第一组第一频谱部分进行编码,其中,所述第一组第一频谱部分包括所述预定源区域以及与所述预定源区域不同的频谱部分,

其中,所述参数计算器被配置成:当目标铺片大于源铺片时,将所述源铺片多次拷贝到所述目标铺片中直到所述目标铺片被所述源铺片的多个拷贝填满为止,或者其中,所述参数计算器(1134)被配置用于:在执行相关性处理以获得匹配标识之前,在频谱上白化(1142)第一频谱部分和第二频谱部分的所述不同对中的所述第一频谱部分或所述第二频谱部分,其中,所述在频谱上白化被执行以使得粗糙的包络信息被去除并且频谱精细结构被强调,或者其中,在频谱上白化(1142)包括计算所述第一频谱部分或所述第二频谱部分的频谱包络估计以及将所述第一频谱部分或所述第二频谱部分的频谱分别除以由所述频谱包络估计指示的频谱包络。

8.根据权利要求7所述的设备,

其中,所述参数计算器(1134)被配置用于:使用所述第二组第二频谱部分中的预定目标区域或者所述第一组第一频谱部分中的预定源区域。

9.根据权利要求7所述的设备,

其中,所述参数计算器(1134)被配置成使得:所述预定目标区域是非交叠的,或者所述预定源区域是交叠的,或者其中,所述预定源区域是所述第一组第一频谱部分的低于间隙填充起始频率(309)的子集,或者其中,预定目标区域覆盖与所述间隙填充起始频率(309)重合的最低频谱区域。

10.根据权利要求7所述的设备,

其中,所述参数计算器(1134)被配置用于:比较目标区域和源区域的对以及所述目标区域和同一源区域的对,其中,所述同一源区域被移位了相关性滞后以提供关于所选择的对的所述相关性滞后的信息作为额外的匹配信息。

11.根据权利要求7所述的设备,

其中,所述参数计算器(1134)被配置用于:执行相关性处理以获得所述第一频谱部分

和所述第二频谱部分的对的匹配结果,所述匹配结果具有负符号,并且

其中,所述参数计算器(1134)被配置成:提供关于所述负符号的信息作为额外的匹配信息。

12.根据权利要求7所述的设备,

其中,所述参数计算器(1134)被配置成:确定整数个目标铺片并且确定用于每个目标铺片的多个相等尺寸的源铺片。

13.根据权利要求7所述的设备,

其中,所述参数计算器(1134)被配置用于:使用以下过程中的至少一个过程来计算所述第一频谱部分或所述第二频谱部分的频谱包络:

利用离散余弦变换DCT来变换所述频谱,通过将较高DCT系数设置为零来保留较低频率DCT系数,并且计算逆DCT;

计算在时域音频帧上计算出的一组线性预测系数的频谱包络;或者

利用低通滤波器对修正离散余弦变换频谱进行滤波。

14.根据权利要求7所述的设备,还包括:源铺片修剪操作(1146)以及用于存储先于当前帧的较早帧的源铺片信息的存储器(1148)。

15.根据权利要求7所述的设备,

其中,源修剪操作(1146)包括:关于多个源铺片的相似度而分析所述多个源铺片,并且从用于互相关性计算(1140)的一组潜在铺片中去除与不同源铺片的相似度大于预定阈值的源铺片。

16.根据权利要求11所述的设备,

其中,所述参数计算器被配置用于:当当前帧中的源区域都不与所述当前帧的目标区域相关、相对于前一帧好于预定阈值时,保留来自所述前一帧的用于每个目标区域的一组匹配信息。

17.一种对编码信号进行解码的方法,包括:

对第一组第一频谱部分的编码表示进行解码(1102)以获得经解码的第一组第一频谱部分(1101),

对第二组第二频谱部分的编码参数化表示进行解码(1104)以获得所述参数化表示的解码表示(1103),其中,所述参数化表示的所述解码表示(1103)对于每个目标频率铺片而言包括源区域标识作为匹配信息;以及

使用由所述匹配信息标识的来自所述第一组第一频谱部分(1101)的源区域来再生(1106)目标频率铺片,

其中,所述再生(1106)包括:使用可控制的白化滤波器(1122,1123),

其中,所述参数化表示的所述解码表示(1103)包括白化信息(1123),

其中,所述再生(1106)包括:当根据所述匹配信息选择的源区域的所述白化信息指示所述源区域要被白化时,在执行频谱包络调整(1128)之前将所述白化滤波器(1122)应用于所述源区域,并且

其中,应用所述白化滤波器被执行以使得粗糙的包络信息被去除并且频谱精细结构被强调,或者其中,应用所述白化滤波器包括计算频谱包络估计以及将所述源区域的频谱除以由所述频谱包络估计指示的频谱包络。

18.一种对音频信号进行编码的方法,包括:

将音频信号转换(1130)成频谱表示;

分析(1132)所述频谱表示以确定要以第一频谱分辨率编码的第一组第一频谱部分以及要以第二频谱分辨率编码的第二组第二频谱部分,其中,所述第二频谱分辨率低于所述第一频谱分辨率;

计算(1134)预定源区域与目标区域之间的相似度,源区域包括频谱部分,且目标区域包括第二频谱部分,

其中,所述计算(1134)包括:比较第一频谱部分和第二频谱部分的不同对的匹配结果以确定所选择的匹配对,以及提供标识所述匹配对的匹配信息;以及

对所述第一组第一频谱部分进行编码(1136),其中,所述第一组第一频谱部分包括所述预定源区域以及与所述预定源区域不同的频谱部分,

其中,所述计算(1134)包括:当目标铺片大于源铺片时,将所述源铺片多次拷贝到所述目标铺片中直到所述目标铺片被所述源铺片的多个拷贝填满为止,或者其中,所述计算(1134)包括:在执行相关性处理以获得匹配标识之前,在频谱上白化(1142)第一频谱部分和第二频谱部分的所述不同对中的所述第一频谱部分或所述第二频谱部分,其中,所述在频谱上白化被执行以使得粗糙的包络信息被去除并且频谱精细结构被强调,或者其中,在频谱上白化(1142)包括计算所述第一频谱部分或所述第二频谱部分的频谱包络估计以及将所述第一频谱部分或所述第二频谱部分的频谱分别除以由所述频谱包络估计指示的频谱包络。

19.一种其上存储有计算机程序的存储介质,所述计算机程序用于当在计算机或处理器上运行时执行根据权利要求17或18所述的方法中的任何一种方法。

用自适应频谱铺片选择的音频信号解码和编码设备及方法

技术领域

[0001] 本发明涉及音频编码/解码,并且特别地涉及使用智能间隙填充的音频编码。

背景技术

[0002] 音频编码为使用心理声学知识处理音频信号中的冗余及无关部分的信号压缩领域。如今音频编解码器通常需要大约60kbps/通道以对几乎任何类型的音频信号进行感知透明编码。较新的编解码器旨在使用如带宽扩展(BWE)的技术通过利用信号中的频谱相似度来降低编码比特率。BWE方案使用低比特率参数集来表示音频信号的高频(HF)分量。HF频谱被来自低频(LF)区域的频谱内容填满,并且频谱形状、倾斜和时间连续性被调整以维持原始信号的音质和音色。这样的BWE方法使得音频编解码器即使在大约24kbps/通道的低比特率下仍能够保持良好的品质。

[0003] 音频信号的存储或传输经常受到严格的比特率约束。在过去,仅当有很低的比特率可用时编码器才被迫大幅度地减小传输音频带宽。

[0004] 现代音频编解码器如今能够通过使用带宽扩展(BWE)方法[1]对宽带信号进行编码。这些算法依赖于高频内容(HF)的参数化表示以及参数驱动的后处理的应用,高频内容(HF)的参数化表示借助于将解码信号的波形编码的低频部分(LF)转移至HF频谱区域中(“修补(patching)”)来生成。按照BWE方案,高于给定的所谓的交叉频率的HF频谱区域的重构通常基于频谱修补。通常,HF区域包括多个相邻补片,并且这些补片中的每一者源自低于给定的交叉频率的LF频谱的带通(BP)区域。现有技术系统通过将一组相邻子带系数从源区域拷贝至目标区域来有效地执行滤波器组表示例如正交镜像滤波器组(QMF)内的修补。

[0005] 在如今的音频编解码器中发现的提高压缩效率从而使得能够在低比特率下扩展音频带宽的另一种技术是音频频谱的适当部分的参数驱动合成替换。例如,原始音频信号的类噪声信号部分可以用在解码器中生成并且由边信息参数缩放的人造噪声来替换,而基本上不损失主观品质。一个示例是包括在MPEG-4高级音频编码(AAC)中的感知噪声替代(PNS)工具[5]。

[0006] 还提供了也使得能够在低比特率下扩展带宽的另一种技术,其是包括在MPEG-D统一语音与音频编码(USAC)[7]中的噪声填充技术。通过由于太粗糙的量化而由量化器的死区得出的频谱间隙(零)随后以解码器中的人工噪声来填充并且通过参数驱动后处理来缩放。

[0007] 另一种现有技术系统被称为精确频谱替换(ASR)[2-4]。除了波形编解码器以外,ASR还采用专用信号合成段,其在解码器处恢复信号的感知上重要的正弦部分。此外,[5]中描述的系统依赖于波形编码器的HF区域中的正弦建模以使得扩展的音频带宽在低比特率下仍具有还不错的感知品质。所有这些方法涉及将数据变换至除了修正离散余弦变换(MDCT)以外的第二域的变换并且还涉及用于保留HF正弦分量的相当复杂的分析/合成段。

[0008] 图13A示出了用于如在例如高效高级音频编码(HE-AAC)中使用的带宽扩展技术的音频编码器的示意图。线1300处的音频信号被输入至包括低通1302和高通1304的滤波器系

统。通过高通滤波器1304输出的信号被输入至参数提取器/编码器1306。例如,参数提取器/编码器1306被配置成用于对参数例如频谱包络参数、噪声添加参数、丢失的谐波参数或逆滤波参数进行计算和编码。所提取的这些参数被输入至比特流复用器1308。低通输出信号被输入至通常包括下采样器1310和核心编码器1312的功能的处理器。低通1302将要被编码的带宽限于与出现在线1300上的原始输入音频信号中出现的带宽相比明显较小的带宽。这由于以下事实而提供了明显的编码增益:核心编码器中出现的全部功能仅必须对具有减小了的带宽的信号进行操作。当例如线1300上的音频信号的带宽为20kHz时并且当低通滤波器1302具有例如4kHz的带宽时,为了满足采样定理,理论上充分的是,下采样器之后的信号的采样频率为8kHz,其基本上降低至必须至少是40kHz的音频信号1300所需的采样速率。

[0009] 图13B示出了相应的带宽扩展解码器的示意图。解码器包括比特流复用器1320。比特流解复用器1320提取核心解码器1322的输入信号和参数解码器1324的输入信号。在上面的示例中,核心解码器输出信号的采样速率为8kHz并且因此带宽为4kHz,然而,为了完成带宽重构,高频重构器1330的输出信号必须处于20kHz,其要求至少40kHz的采样速率。为了使这可能,需要具有上采样器1325和滤波器组1326的功能的解码器处理器。然后,高频重构器1330接收由滤波器组1326输出的经频率分析的低频信号,并且使用高频带的参数化表示对由图13A的高通滤波器1304限定的频率范围进行重构。高频重构器1330具有几个功能如使用低频范围中的源范围来再生上频率范围、频谱包络调整、噪声添加功能以及在上频率范围内引入丢失的谐波的功能,并且为了解释较高频率范围通常与较低频率范围具有不同的音调的事实,如果在图13A的编码器中施加以及计算,则还包括逆滤波操作。在HE-ACC中,丢失的谐波在解码器侧被重新合成,并且被精确地置于重构频带的中间。因此,在某个重构频带中确定了的所有丢失的谐波线未被放置在其位于原始信号中的频率值处。替代地,那些丢失的谐波线被置于某个频带的中心处的频率处。从而,当原始信号中的丢失的谐波线被放置得很靠近原始信号中的重构频带边界时,通过将经重构的信号中的该丢失的谐波线置于频带的中心处而引入的频率的误差接近生成并且发送了参数的个别重构频带的50%。

[0010] 此外,尽管典型的音频核心编码器操作在频谱域中,然而核心解码器生成时域信号,然后该时域信号通过滤波器组1326功能再次被转换至频谱域。这引入了额外的处理延迟,由于首先从频谱域变换成频域以及再次变换成通常不同的频域的串联处理而可能引入伪声,并且当然这还要求大量的计算复杂度以及因此电功率,当带宽扩展技术被应用于移动装置如移动电话、平板或手提计算机等时,电力尤其是问题。

[0011] 当前音频编解码器使用BWE作为编码方案的主要部分来执行低比特率音频编码。然而,BWE技术限于仅替换高频(HF)内容。此外,它们不能使得高于给定的交叉频率的感知上重要的内容能够被波形编码。因此,由于在大多数系统中不考虑信号的音调谐波的精确对准,所以当实现BWE时,当代音频编解码器丢失HF细节或音品。

[0012] 目前现有技术的BWE系统的另一缺点是需要将音频信号变换至新的域中以实现BWE(例如,从MDCT域变换至QMF域)。这导致同步复杂、额外的计算复杂度以及增加了的存储需求。

[0013] 通常,带宽扩展方案使用频谱修补以用于重构高于给定的所谓交叉频率的高频频谱区的目的。HF区包括多个相邻补片,并且这些补片中的每个源自低于给定的交叉频率的低频频谱的同一带通区域。在信号的滤波器组表示内,这样的系统将低频频谱中的一组相

邻的子带系数复制到HF区中。所选组的边界通常与系统有关,而与信号无关。对于一些信号内容,这种静态修补选择可能会导致经重构的信号的不好的音色及着色。

[0014] 其他方法通过信号自适应单边频带(SSB)调制将LF信号传送至HF区。与复制过程相比,这样的方法计算复杂度高,这是因为其以高采样速率对时域信号进行操作。

[0015] 此外,修补可能不稳定,特别是对于非音调信号如无声语音更是如此。因此,已知的修补方案可能对音频信号带来损伤。

发明内容

[0016] 本发明的目的是提供一种得到较好音频质量的改进的编码/解码构思。

[0017] 该目的通过解码设备、编码设备、音频解码方法或其上存储有计算机程序的存储介质来实现。

[0018] 又一方面基于以下发现:可以通过应用信号自适应频率铺片填充方案来弥补音频品质的损伤。为此,在编码器侧执行分析以找出某个目标区域的最匹配的候选源区域。针对目标区域标识某个源区域的匹配信息连同可选地一些附加信息一起被生成,并且作为边信息被发送至解码器。然后,解码器使用该匹配信息来应用频率铺片填充操作。为此,解码器从发送的数据流或数据文件读取匹配信息,并且访问针对某个重构频带而识别的源区域,并且如果在匹配信息中指示,则额外地执行该源区域数据的某种处理以生成重构频带的原始频谱数据。然后,使用频谱包络信息对该频率铺片填充操作的结果即重构频带的原始频谱数据进行整形,以最终获得包括第一频谱部分如音调部分的重构频带。然而,这些音调部分不是通过自适应铺片填充方案生成,但是这些第一频谱部分由音频解码器或核心解码器直接输出。

[0019] 自适应频谱铺片选择方案可以以低粒度来操作。在该实现中,源区域被细分成通常交叠的源区域,并且目标区域或重构频带通过非交叠的频率目标区域来给出。然后,在编码器侧确定每个源区域与每个目标区域之间的相似度,并且通过匹配信息来标识源区域与目标区域的最佳匹配对,并且在解码器侧,在匹配信息中标识的源区域用于生成重构频带的原始频谱数据。

[0020] 出于获得较高粒度的目的,允许每个源区域移位以获得其中相似度最大的某个滞后。该滞后可以如区元一样精细,并且允许源区域与目标区域之间的甚至更佳的匹配。

[0021] 此外,除了仅标识最佳匹配对以外,该相关性(correlation)滞后还可以在匹配信息内被发送,并且另外地,甚至可以发送符号。当在编码器侧确定该符号为负时,则还在匹配信息内发送相应的符号旗标,而在解码器侧,将源区域频谱值乘以“-1”,或者在复数表示中,将源区域频谱值“旋转”180°。

[0022] 本发明的又一实现应用铺片白化操作。频谱的白化去除粗糙的频谱包络信息,并且强调对于评估铺片相似度而言最感兴趣的频谱精细结构。因此,在计算互相关性量度之前,一方面对频率铺片和/或另一方面对源信号进行白化。当使用预定的过程仅对铺片进行白化时,发送向解码器表示应当对IGF内的频率应用相同预定的白化处理的白化旗标。

[0023] 关于铺片选择,优选地使用相关性的滞后来将再生的频谱在频谱上移位整数个变换区元。取决于基本的变换,频谱移位可能需要加法相关。在奇数滞后的情况下,该铺片额外地通过乘以交替的时间序列-1/1来被调制以补偿MDCT内的每个其他频带的频率颠倒表

示。此外,当生成频率铺片时,应用相关性结果的符号。

[0024] 此外,优选地使用铺片修剪和稳定化,以确保避免由同一重构区域或目标区域的快速变化的源区域导致的伪声。为此,不同的所标识的源区域之间的相似度分析被执行,并且当源铺片以高于阈值的相似度类似于其他源铺片时,则可以从一组潜在的源铺片中丢弃该源铺片,因为与其他源铺片高度相关。此外,作为一种铺片选择稳定化,如果当前帧中的源铺片都不与当前帧中的目标铺片相关(好于给定阈值),则优选地保持来自前一帧的铺片顺序。

[0025] 另一方面基于以下发现:可以通过IGF提高经重构的信号音频质量,这是由于整个频谱可由核心编码器访问,使得例如高频谱范围中的感知上重要的音调部分仍可以通过核心编码器编码而不是进行参数化替代。另外,执行使用来自第一组第一频谱部分的频率铺片的间隙填充操作,该第一组第一频谱部分例如为通常来自较低频率范围但是在可用的情况下也可以来自较高频率范围的一组音调部分。然而对于在编码器侧执行频谱包络调整,来自位于重构频带中的第一组频谱部分的频谱部分不通过例如频谱包络调整进行进一步的后处理。仅重构频带中的不源自于核心编码器的其余频谱值要使用包络信息来包络调整。优选地,包络信息是将重构频带中的第一组第一频谱部分和同一重构频带中的第二组第二频谱部分的能量计入在内的全频带包络信息,其中,第二组第二频谱部分中的靠后的频谱值被指示为零,因此,不通过核心编码器编码,而是使用低分辨率能量信息来参数化编码。

[0026] 已发现,不论是关于相应频带的带宽而归一化的绝对能量值还是未归一化的绝对能量值对于解码器侧的应用是有用的并且非常高效。这尤其适用于当增益因子必须基于重构频带中的残差能量、重构频带中的丢失能量及重构频带中的频率铺片信息来计算的情况。

[0027] 此外,优选的是,编码比特流不仅包括重构频带的能量信息,而且另外包括延伸直至最大频率的比例因子频带的比例因子。这确保了针对特定音调部分即第一频谱部分可用的每个重构频带,该第一组第一频谱部分可以实际上以合适的幅度来解码。此外,除了每个重构频带的比例因子之外,在编码器中还生成该重构频带的能量并且将其传送至解码器。此外,优选的是,重构频带与比例因子频带重合,或者在能量分组的情况下,至少重构频带的边界与比例因子频带的边界重合。

[0028] 另一方面基于以下发现:可以通过在核心解码器所操作的相同频谱域中执行带宽扩展来解决和克服一方面与带宽扩展的分离有关而另一方面与核心编码有关的问题。因此,提供对全音频信号范围进行编码和解码的全速率核心解码器。这不要求在编码器侧需要下采样器以及在解码器侧要求上采样器。替代地,在全采样速率或全带宽域中执行全部处理。为了获得高编码增益,分析音频信号以找到必须要以高分辨率编码的第一组第一频谱部分,其中,在一种实施方式中,第一组第一频谱部分可以包括音频信号的音调部分。另一方面,音频信号中的构成第二组第二频谱部分的非音调或有噪声的分量以低频谱分辨率参数化地被编码。编码音频信号则仅需要以高频谱分辨率按照波形保留方式被编码的第一组第一频谱部分以及额外地使用源自第一组的频率“铺片”以低分辨率参数化地被编码的第二组第二频谱部分。在解码器侧,作为全频带解码器的核心解码器按照波形保留方式即无需知道存在任何额外的频率再生来重构第一组第一频谱部分。然而,如此生成的频谱具

有很多频谱间隙。这些间隙随后通过一方面使用应用参数化数据的频率再生以及另一方面使用源频谱范围即由全速率音频解码器重构的第一频谱部分而用本发明的智能间隙填充 (IGF) 技术来填充。

[0029] 在另外的实施方式中,仅通过噪声填充而非带宽复制或频率铺片填充而重构的频谱部分包括第三组第三频谱部分。由于本编码构思一方面在核心编码/解码的单个域中操作而另一方面在频率再生的单个域中操作的事实,通过不使用频率再生的噪声填充或者通过使用不同的频率范围处的频谱铺片的频率再生,IGF不仅仅被限于填充较高的频率范围,而且还可以填充较低的频率范围。

[0030] 此外,强调的是,关于频谱能量的信息、关于个体能量的信息或者个体能量信息、关于残留能量的信息或残留能量信息、关于铺片能量的信息或铺片能量信息、或关于丢失能量的信息或丢失能量信息可以不仅包括能量值,而且还可以包括(例如,绝对)幅度值、电平值或任何其他值,根据这些可以得到最终能量值。因此,关于能量的信息可以包括例如能量值本身和/或电平的值和/或幅度的值和/或绝对幅度的值。

[0031] 又一方面基于以下发现:相关性情况不仅对于源范围重要,而且对于目标范围也重要。此外,本发明承认以下情况:在源范围和目标范围中可以出现不同的相关性情况。例如,当考虑具有高频噪声的语音信号时,该情况可以是:当扬声器被放置在中间时,包括具有小数目的泛音的话音信号的低频频带在左通道和右通道中高度相关。然而,由于与右侧的其他高频噪声或无高频噪声相比在左侧可能存在不同的高频噪声的事实,高频部分可以强烈地不相关。从而,当可以执行忽略该情况的简单的间隙填充操作时,则高频部分也会相关,并且这可以在经重构的信号中生成严重的空间隔离伪声。为了解决这个问题,计算重构频带的参数化数据或通常必须要使用第一组第一频谱部分被重构的第二组第二频谱部分的参数化数据以标识第二频谱部分或者不同地陈述的重构频带的第一双通道表示或不同的第二双通道表示。因此,在编码器侧,针对第二频谱部分计算双通道标识,即,针对额外地计算重构频带的能量信息的部分计算双通道标识。然后,解码器侧的频率再生器根据第一组第一频谱部分即源范围中的第一部分以及第二部分的参数化数据如频谱包络能量信息或任何其他频谱包络数据以及额外地根据第二部分的双通道标识即重新考虑的情况下的该重构频带的双通道标识来再生第二频谱部分。

[0032] 优选地,双通道标识作为每个重构频带的旗标被发送,并且该数据被从编码器发送至解码器,然后解码器按照由优选地计算出的关于核心频带的旗标所指示的对核心信号进行解码。于是,在一种实现中,核心信号以两种立体声表示(例如,左/右和中间/侧)被存储,并且对于IGF频率铺片填充,按照由对于智能间隙填充或重构频带即对于目标范围的双通道标识旗标所指示的,选择源铺片表示来匹配目标铺片表示。

[0033] 强调的是,该过程不仅对立体声信号即对左通道和右通道起作用,而且对于多通道信号也工作。在多通道信号的情况下,可以以该方式对几对不同通道如左通道和右通道作为第一对,左环绕通道和右环绕通道作为第二对以及中心通道和LFE通道作为第三对进行处理。对于较高的输出通道格式如7.1、11.1等,可以确定其他对。

[0034] 另一方面基于以下发现:通过将时间噪声整形(TNS)或时间铺片整形(TTS)技术与高频重构相组合来获得改进的质量和降低的比特率,特别是对于包括在音频信号中经常出现的瞬态部分的信号而言,提高了品质并且减小了比特率。编码器侧的通过关于频率的预

测来实现的TNS/TTS处理重构了音频信号的时间包络。取决于该实现,即,当在不仅包括源频率范围而且还包括要在频率再生解码器中重构的目标频率范围的频率范围内确定时间噪声整形滤波器时,时间包络不仅被应用于直至间隙填充起始频率的核心音频信号而且还被应用于经重构的第二频谱部分的频谱范围。从而,在不使用时间铺片整形的情况下出现的预回声和后回声被减少或消除。这通过不仅在直至某个间隙填充起始频率的核心频率范围内而且在高于核心频率范围的频率范围内应用关于频率的逆预测来完成。为此,在应用关于频率的预测之前,在解码器侧执行频率再生或频率铺片生成。然而,取决于是对滤波之后的频谱残差值还是对包络整形之前的(全)频谱值执行能量信息计算,可以在频谱包络整形之前或频谱包络整形之后应用关于频率的预测。

[0035] 关于一个或更多个频率铺片的TTS处理额外地建立源范围与重构范围之间的或两个相邻的重构范围或频率铺片内的相关性的连续性。

[0036] 按照一种实现,优选地使用复数TNS/TTS滤波。从而,避免临界采样的实数表示如MDCT的(时间)混叠伪声。通过不仅应用修正离散余弦变换还应用修正离散正弦变换,除了获得复数修正变换以外还能够在编码器侧计算复数TNS滤波器。然而,仅修正离散余弦变换值即复数变换的实部被发送。然而,在解码器侧,可以使用先前或后续帧的MDCT频谱来估计变换的虚部以使得在解码器侧可以在关于频率的逆预测中特别是在关于源范围与重构范围之间的边界的预测以及还有关于重构范围内的频率相邻的频率铺片之间的边界的预测中,再次应用复数滤波器。

[0037] 本发明的音频编码系统有效地以宽范围的比特率对任意音频信号进行编码。然而,对于高比特率,本发明的系统会聚于透明度,对于低比特率,使感知烦恼最小化。因此,可用比特率的主要份额用于在编码器中对信号的仅感知上最相关的结构进行波形编码,并且结果的频谱间隙在解码器中以大致近似原始频谱的信号内容来填充。非常有限的比特预算被消耗以通过从编码器发送至解码器的专用边信息来控制参数驱动的所谓的频谱智能间隙填充(IGF)。

附图说明

[0038] 随后,参照附图描述本发明的优选实施方式,在附图中:

[0039] 图1A示出了用于对音频信号进行编码的设备;

[0040] 图1B示出了与图1A的编码器匹配的用于对编码音频信号进行解码的解码器;

[0041] 图2A示了解码器的优选实现;

[0042] 图2B示出了编码器的优选实现;

[0043] 图3A示出了由图1B的频谱域解码器生成的频谱的示意性表示;

[0044] 图3B示出了表示比例因子频带的比例因子与用于重构频带的能量之间的关系及噪声填充频带的噪声填充信息的表格;

[0045] 图4A示出了频谱域编码器的用于将频谱部分的选择应用至第一组频谱部分和第二组频谱部分的功能;

[0046] 图4B示出了图4A的功能的实现;

[0047] 图5A示出了MDCT编码器的功能;

[0048] 图5B示出了使用MDCT技术的解码器的功能;

- [0049] 图5C示出了频率再生器的实现；
- [0050] 图6A示出了具有时间噪声整形/时间铺片整形功能的音频编码器；
- [0051] 图6B示出了使用时间噪声整形/时间铺片整形技术的解码器；
- [0052] 图6C示出了时间噪声整形/时间铺片整形功能的具有频谱预测滤波器和频谱整形器的不同顺序的又一功能；
- [0053] 图7A示出了时间铺片整形 (TTS) 功能的实现；
- [0054] 图7B示出了与图7A的编码器实现相匹配的解码器实现；
- [0055] 图7C示出了原始信号和不具有TTS的扩展信号的频谱表示；
- [0056] 图7D示出了示出智能间隙填充频率与时间铺片整形能量之间的对应关系的频率表示；
- [0057] 图7E示出了原始信号和具有TTS的扩展信号的频谱图；
- [0058] 图8A示出了具有频率再生的双通道解码器；
- [0059] 图8B示出了示出各表示和源/目的范围的不同组合的表格；
- [0060] 图8C示出了示出具有图8A的频率再生的双通道解码器的功能的流程图；
- [0061] 图8D示出了图8A的解码器的更详细的实现；
- [0062] 图8E示出了用于要由图8A的解码器解码的双通道处理的编码器的实现；
- [0063] 图9A示出了具有使用再生频率范围的能量值的频率再生技术的解码器；
- [0064] 图9B示出了图9A的频率再生器的更详细的实现；
- [0065] 图9C示出了示出图9B的功能的示意图；
- [0066] 图9D示出了图9B的解码器的又一实现；
- [0067] 图10A示出了与图9A的解码器匹配的编码器的框图；
- [0068] 图10B示出了用于说明图10A的参数计算器的又一功能的框图；
- [0069] 图10C示出了示出图10A的参数化计算器的又一功能的框图；
- [0070] 图10D示出了图10A的参数化计算器的又一功能的框图；
- [0071] 图11A示出了具有解码器中的频谱加铺片操作的特定源范围标识的又一解码器；
- [0072] 图11B示出了图11A的频率再生器的又一功能；
- [0073] 图11C示出了用于与图11A中的解码器协作的编码器；
- [0074] 图11D示出了图11C的参数计算器的实现的框图；
- [0075] 图12A和图12B示出了用于示出源范围和目标范围的频率示意图；
- [0076] 图12C示出了两个信号的示例相关性的图；
- [0077] 图13A示出了具有带宽扩展的现有技术编码器；以及
- [0078] 图13B示出了具有带宽扩展的现有技术解码器。

具体实施方式

[0079] 图1A示出了用于对音频信号99进行编码的设备。音频信号99被输入至时间频谱转换器100,时间频谱转换器100用于将具有采样速率的音频信号转换成由该时间频谱转换器输出的频谱表示101。频谱101被输入至用于分析频谱表示101的频谱分析器102。频谱分析器101被配置成确定要以第一频谱分辨率被编码的第一组第一频谱部分103以及要以第二频谱分辨率被编码的不同的第二组第二频谱部分105。第二频谱分辨率小于第一频谱分辨

率。第二组第二频谱部分105被输入至用于计算具有第二频谱分辨率的频谱包络信息的参数计算器或参数化编码器104。此外,设置频谱域音频编码器106用于生成具有第一频谱分辨率的第一组第一频谱部分的第一编码表示107。此外,参数计算器/参数化编码器104被配置成用于生成第二组第二频谱部分的第二编码表示109。第一编码表示107和第二编码表示109被输入至比特流复用器或比特流形成器108,并且块108最终输出编码音频信号以进行传输或存储在存储装置上。

[0080] 通常,第一频谱部分如图3A的306会由两个第二频谱部分如307A和307B围绕。在HE AAC中并不是这种情况,在HE AAC中核心编码器频率范围频带受限。

[0081] 图1B示出了与图1A的编码器匹配的解码器。第一编码表示107被输入至用于生成第一组第一频谱部分的第一解码表示的频谱域音频解码器112,该解码表示具有第一频谱分辨率。此外,第二编码表示109被输入至用于生成具有低于第一频谱分辨率的第二频谱分辨率的第二组第二频谱部分的第二解码表示的参数化解码器114。

[0082] 解码器还包括用于使用第一频谱部分再生具有第一频谱分辨率的经重构的第二频谱部分的频率再生器116。频率再生器116执行铺片(tile)填充操作,即,使用第一组第一频谱部分的铺片或一部分并且将第一组第一频谱部分拷贝至具有第二频谱部分的重构范围或重构频带中,并且通常执行频谱包络整形或由参数化解码器114所输出的第二解码表示所指示的其他操作,即,通过使用关于第二组第二频谱部分的信息。如在线117上的频率再生器116的输出端处所示,经解码的第一组第一频谱部分和经重构的第二组频谱部分被输入至频谱时间转换器118,频谱时间转换器118被配置成用于将第一解码表示和经重构的第二频谱部分转换成时间表示119,该时间表示具有某个高的采样速率。

[0083] 图2B示出了图1A的编码器的实现。音频输入信号99被输入至与图1A的时间频谱转换器100对应的分析滤波器组220。然后,在TNS块222中执行时间噪声整形操作。因此,至与图2B的音调掩蔽块226对应的图1A的频谱分析器102的输入在当时间噪声整形/时间铺片整形操作不被应用时可以是全部频谱值,并且在当如2B所示的块222的TNS操作被应用时可以是频谱残差值。对于双通道信号或多通道信号,可以额外地执行联合通道编码228以使得图1A的频谱域编码器106可以包括联合通道编码块228。此外,提供用于执行无损数据压缩的熵编码器232,其也是图1A的频谱域编码器106的一部分。

[0084] 频谱分析器/音调掩蔽226将TNS块222的输出分离成与第一组第一频谱部分103对应的核心频带和音调分量以及图1A的第二组第二频谱部分105对应的残差分量。被表示为IGF参数提取编码的块224对应于图1A的参数化编码器104,并且比特流复用器230对应于图1A的比特流复用器108。

[0085] 优选地,分析滤波器组222被实现为MDCT(修正离散余弦变换滤波器组),并且该MDCT用于使用用作频率分析工具的修正离散余弦变换将信号99变换至时间频率域。

[0086] 优选地,频谱分析器226应用音调掩蔽。该音调掩蔽估计段用于将信号中的音调分量与类噪声分量分离开。这使得核心编码器228能够使用心理声学模块对所有音调分量进行编码。音调掩蔽估计段可以按照大量不同的方式来实现,并且优选地被实现成在功能方面类似于在语音/音频编码的正余弦建模中使用的正弦轨迹估计段[8,9]或在[10]中描述的基于HILN模型的音频编码器。优选地,使用一种易于实现而不必须维持生死轨迹的实现,但是还可以使用任何其他音调或噪声检测器。

[0087] IGF模块计算源区域与目标区域之间存在的相似度。目标区域将由来自源区域的频谱表示。使用互相关性方法来进行源区域与目标区域之间的相似度的量度。目标区域被划分成 n_{tar} 个非交叠频率铺片。对于目标区域中的每个铺片,从固定起始频率起创建 n_{src} 个源铺片。这些源铺片以0与1之间的因子交叠,其中0表示0%交叠,1表示100%交叠。这些源铺片中的每一个与各种滞后处的目标铺片相关,以找出与目标铺片最佳地匹配的源铺片。最佳匹配铺片数目被存储在 $tileNum[idx_{tar}]$ 中,在其处与目标最相关的滞后被存储在 $xcorr_lag[idx_{tar}][idx_{src}]$ 中,并且相关性的符号被存储在 $xcorr_sign[idx_{tar}][idx_{src}]$ 中。在高度负相关的情况下,源铺片需要在解码器处进行铺片填充处理之前乘以-1。由于使用音调掩蔽来保存音调分量,所以IGF模块还要注意不覆写频谱中的音调分量。带状能量参数用于存储目标区域的能量,其使得能够准确地重构频谱。

[0088] 该方法相对于经典的SBR[1]具有的某些优点在于:核心编码器保存了多音调信号的谐波网格,而仅正弦波之间的间隙填充有来自源区域的最匹配的“整形噪声”。该系统相比ASR(精确频谱替换)[2-4]而言的另一优点是不存在信号合成段,该信号合成段在解码器处创建信号的重要部分。替代地,该任务由核心编码器接管,使得能够保存频谱的重要分量。所提出的系统的另一优点是特征所提供的连续的可伸缩性。对于每个铺片仅使用 $tileNum[idx_{tar}]$ 以及 $xcorr_lag=0$,其被称为粗粒度匹配,并且当对于每个铺片使用变量 $xcorr_lag$ 时,可以用于低比特率,使得能够较佳地匹配目标频谱和源频谱。

[0089] 另外,提出了铺片选择稳定化技术,其去除频域伪声如颤音噪声或音乐噪声。

[0090] 在立体声通道对的情况下,应用额外的联合立体声处理。这是必要的,因为对于某个目的范围,信号可以来自高度相关的平移声音源。在针对该特定区域而选择的源区域非良好相关的情况下,虽然能量针对目的区域被匹配,但空间图像可能由于不相关的源区域而受损。编码器分析每个目的区域能量频带,通常执行频谱值的互相关性,并且如果超过某个阈值,则为该能量频带设置联合旗标。在解码器中,如果未设置联合立体声旗标,则单独地处理左通道能量频带和右通道能量频带。在设置联合立体声旗标的情况下,在联合立体声域执行能量和修补两者。IGF区域的联合立体声信息类似于核心编码的联合立体声信息被用信号发送,包括在预测情况下指示预测的方向是从下混合至残差还是从残差至下混合的旗标。

[0091] 能量可以根据L/R域中的传送能量来计算。

[0092] $midNrg[k] = leftNrg[k] + rightNrg[k];$

[0093] $sideNrg[k] = leftNrg[k] - rightNrg[k];$

[0094] 其中, k 是变换域中的频率索引。

[0095] 另一解决方案是在联合立体声域中针对其中联合立体声是活动的频带直接计算和发送能量,所以在解码器侧不需要额外的能量变换。

[0096] 源铺片总是根据中间/侧矩阵来创建:

[0097] $midTile[k] = 0.5 \cdot (leftTile[k] + rightTile[k])$

[0098] $sideTile[k] = 0.5 \cdot (leftTile[k] - rightTile[k])$

[0099] 能量调整:

[0100] $midTile[k] = midTile[k] * midNrg[k];$

[0101] $sideTile[k] = sideTile[k] * sideNrg[k];$

[0102] 联合立体声->LR变换:

[0103] 如果没有额外的参数被编码:

[0104] $\text{leftTile}[k] = \text{midTile}[k] + \text{sideTile}[k]$

[0105] $\text{rightTile}[k] = \text{midTile}[k] - \text{sideTile}[k]$

[0106] 如果有额外的预测参数被编码并且如果用信号发送的方向是从中间至侧:

[0107] $\text{sideTile}[k] = \text{sideTile}[k] - \text{predictionCoeff} \cdot \text{midTile}[k]$

[0108] $\text{leftTile}[k] = \text{midTile}[k] + \text{sideTile}[k]$

[0109] $\text{rightTile}[k] = \text{midTile}[k] - \text{sideTile}[k]$

[0110] 如果用信号发送的方向是从侧至中间:

[0111] $\text{midTile1}[k] = \text{midTile}[k] - \text{predictionCoeff} \cdot \text{sideTile}[k]$

[0112] $\text{leftTile}[k] = \text{midTile1}[k] - \text{sideTile}[k]$

[0113] $\text{rightTile}[k] = \text{midTile1}[k] + \text{sideTile}[k]$

[0114] 该处理确保:根据用于再生高度相关的目的区域和平移目的区域的铺片,即使源区域不相关,结果的左通道和右通道仍表示相关且平移的声音源,保留了这样的区域的立体声图像。

[0115] 换言之,在比特流中,传送指示应当使用L/R还是M/S——作为一般联合立体声编码的示例——的联合立体声旗标。在解码器中,首先,按照核心频带的联合立体声旗标的指示来解码核心信号。第二,核心信号被存储在L/R表示和M/S表示二者中。对于IGF铺片填充,按照IGF频带的联合立体声信息的指示来选择源铺片表示以匹配目标铺片表示。

[0116] 时间噪声整形(TNS)是标准技术并且是ACC[11-13]的一部分。TNS可以被认为是感知编码器的基本方案的扩展,在滤波器组与量化段之间插入了可选的处理步骤。TNS模块的主要任务是将所产生的量化噪声隐藏在瞬态类似信号的时间掩蔽区域中,从而导致更高效的编码方案。首先,TNS在变换域例如MDCT中使用“前向预测”计算一组预测系数。然后,这些系数用于平坦化信号的时间包络。当量化影响经TNS滤波的频谱时,量化噪声也在时间上平坦。通过在解码器侧应用逆TNS滤波,量化噪声根据TNS滤波器的时间包络被整形,因此量化噪声短暂地被掩蔽。

[0117] IGF基于MDCT表示。为了高效编码,优选地,必须使用大约20ms的长块。如果这样的长块内的信号包括瞬态信号,则由于铺片填充而在IGF频谱带中出现可听见的预回声和后回声。图7C示出了由于IGF的在瞬态开始之前的典型预回声效果。在左侧,示出了原始信号的频谱图,在右侧,示出了没有进行TNS滤波的带宽扩展信号的频谱图。

[0118] 在IGF背景下,通过使用TNS来减小该预回声效果。在此,当在解码器中对TNS残差信号执行频谱再生时,TNS用作时间铺片整形(TTS)工具。照例使用编码器侧的全频谱来计算和应用所需的TTS预测系数。TNS/TTS起始和终止频率不受IGF工具的IGF起始频率 f_{IGFstart} 的影响。与传统的TNS相比,TTS终止频率被增大至IGF工具的终止频率,其高于 f_{IGFstart} 。在解码器侧,再次对全频谱即核心频谱加上再生的频谱加上来自音调图的音调分量(参见图7E)应用TNS/TTS系数。TTS的应用对于以下是必需的:形成再生的频谱的时间包络以再次与原始信号的包络匹配。所以,减小了示出的预回声。另外,仍然照例使用TNS对信号中的低于 f_{IGFstart} 的量化噪声进行整形。

[0119] 在传统的解码器中,对音频信号的频谱修补破坏补片边界处的频谱相关性,并且

因此通过引入分散而损害音频信号的时间包络。因此,对残差信号执行IGF铺片填充的另一益处在于:在应用整形滤波器之后,铺片边界无缝相关,从而导致对信号的更加逼真的时间再现。

[0120] 在本发明的编码器中,经历了TNS/TTS滤波、音调掩蔽处理和IGF参数估计的频谱缺乏除了音调分量以外的任何高于IGF起始频率的信号。该稀疏频谱现在由核心编码器使用算术编码和预测编码的原理来编码。这些编码的分量连同信令比特形成音频的比特流。

[0121] 图2A示出了对应的解码器实现。图2A中的与编码音频信号对应的比特流被输入至解复用器/解码器,解复用器/解码器相对于图1B将被连接至块112和114。比特流解复用器将输入的音频信号分离成图1B的第一编码表示107和图1B的第二编码表示109。具有第一组第一频谱部分的第一编码表示被输入至与图1B的频谱域解码器112对应的联合通道解码块204。第二编码表示被输入至未在图2A中示出的参数化解码器114,然后被输入至与图1B的频率再生器116对应的IGF块202。频率再生所需要的第一组第一频谱部分经由线203被输入至IGF块202。此外,在联合通道解码204之后,在音调掩蔽块206中应用特定核心解码以使得音调掩蔽206的输出对应于频谱域解码器112的输出。然后,通过组合器208执行组合,即,其中组合器208的输出现在具有全范围频谱但仍处于TNS/TTS滤波域的帧构建。然后,在块210中,使用经由线109提供的TNS/TTS滤波器信息执行逆TNS/TTS操作,即,TTS边信息优选地包括在由频谱域编码器106生成的第一编码表示中,或者可以被包括在第二编码表示中,频谱域编码器106可以是例如简单的AAC或USAC核心编码器。在块210的输出处,提供直至最大频率为止的完整频谱,其是由原始输入信号的采样速率限定的全范围频率。然后,在合成滤波器组212中执行频谱/时间转换以最终获得音频输出信号。

[0122] 图3A示出了频谱的示意性表示。该频谱按照比例因子频带SCB被细分,其中,在图3A示出的示例中具有7个比例因子频带SCB1至SCB7。如图3A中示意性地示出的,比例因子频带可以是以AAC标准限定的AAC比例因子频带,并且具有至上频率的增大的带宽。优选的是不从频谱的一开始即在低频处执行智能间隙填充,而是在309处示出的IGF起始频率处开始IGF操作。因此,核心频带从最低频率扩展至IGF起始频率。在IGF起始频率之上,应用频谱分析以将高分辨率频谱分量304、305、306、307(第一组第一频谱部分)与由第二组第二频谱部分表示的低分辨率分量分离开。图3A示出了被示例性地输入至频谱域编码器106或联合通道编码器228的频谱,即,核心编码器在全范围下操作,但是对大量的零频谱值进行编码,即,这些零频谱至被量化成零或者在量化前后被设置成零。无论如何,核心解码器操作在全范围,即,如频谱会如所示出的一样,即,核心解码器不必非得知道具有较低频谱分辨率的第二组第二频谱部分的任何智能间隙填充或编码。

[0123] 优选地,高分辨率由频谱线如MDCT线的线状编码来限定,而第二分辨率或低分辨率通过例如仅计算每比例因子频带的单个频谱值来限定,其中,比例因子频带覆盖几条频率线。从而,第二低分辨率关于其频谱分辨率远低于通常由核心编码器如AAC或USAC核心编码器所应用的线状编码定义的第一分辨率或高分辨率。

[0124] 关于比例因子或能量计算,在图3B中示出了该情况。由于编码器是核心编码器的事实以及由于在每个频带中可以存在但不是必须存在第一组频谱部分的事实,核心编码器不仅在低于IGF起始频率309的核心范围中而且还在高于IGF起始频率直至最大频率 f_{IGFstop} 的范围内计算每个频带的比例因子, f_{IGFstop} 小于或等于采样频率的二分之一即 $f_s/2$ 。从而,

图3a的编码音调部分302、304、305、306、307以及在该实施方式中连同比例因子SCB1至SCB7对应于高分辨率频谱数据。低分辨率频谱数据从IGF起始频率开始被计算,并且对应于能量信息值 E_1 、 E_2 、 E_3 、 E_4 ,其连同比例因子SF4至SF7一起被发送。

[0125] 特别地,当核心编码器处于低比特率条件下时,可以在核心频带即频率低于IGF起始频率,即在比例因子频带SCB1至SCB3中额外地应用额外的噪声填充操作。在噪声填充中,存在被量化成零的几条相邻的频谱线。在解码器侧,这些量化成零的频谱值被重新合成,并且经重新合成的频谱值的量值使用噪声填充能量如图3B中的308处示出的 NF_2 来被调整。尤其关于USAC中的比例因子可以以绝对值或相对值给出的噪声填充能量对应于被量化成零的该组频谱值的能量。这些噪声填充频谱线还可以被视为第三组第三频谱部分,其在没有任何依赖于使用来自其他频率的频率铺片的频率再生的IGF操作的情况下通过简单的噪声填充合成来再生,以用于使用来自源范围的频谱值以及能量信息 E_1 、 E_2 、 E_3 、 E_4 来重构频率铺片。

[0126] 优选地,计算能量信息的频带与比例因子频带重合。在其他实施方式中,应用能量信息值分组以使得例如对于比例因子频带4和5仅发送单个能量信息值,但是甚至在该实施方式中,分组的重构频带的边界与比例因子频带的边界重合。如果应用不同的频带分离,则可以应用某些重新计算或合成,并且这可以根据某个实现而合乎情理。

[0127] 优选地,图1A的频谱域编码器106是如图4A中所示的心理声学驱动的编码器。通常,如按照例如MPEG2/4AAC标准或MPEG1/2,层3标准所示,要在变换成了频谱范围(图4A中的401)之后被编码的音频信号转发至比例因子计算器400。比例因子计算器由心理声学模型控制,该心理声学模型额外地接收要量化的音频信号或如在MPEG1/2层3或按照MPEGAAC标准中接收音频信号的复数频谱表示。心理声学模型针对每个比例因子频带计算用于表示心理声学阈值的比例因子。另外,然后,比例因子通过公知的内外迭代循环的协作或通过任何其他适当的编码过程被调整以使得实现某些比特率条件。然后,一方面要量化的频谱值和另一方面所计算出的比例因子被输入至量化器处理器404。在简单的音频编码器操作中,要量化的频谱值通过比例因子加权,并且然后经加权的频谱值被输入至固定的量化器,该固定的量化器通常具有至上幅度范围的压缩功能。然后,在量化器处理器的输出端处,确实存在量化索引,然后量化索引被转发至熵编码器,熵编码器通常具有对于相邻频率值或在本领域也称作零值的“串(run)”的一组零量化索引的特定并且非常有效的编码。

[0128] 然而,在图1A的音频编码器中,量化器处理器通常接收关于来自频谱分析器的第二频谱部分的信息。从而,量化器处理器404确保:在量化器处理器404的输出中,如由频谱分析器102识别的第二频谱部分为零或者具有由编码器或解码器承认为零表示的表示,零表示可以很高效地被编码,尤其当在频谱中存在零值的“串”时。

[0129] 图4B示出了量化器处理器的实现。MDCT频谱值可以被输入至设置成零块410。然后,在在块412中通过比例因子执行加权之前,第二频谱部分就已经被设置成零。在另一实现中,不提供块410,但是在加权块412之后的块418中执行设置成零协作。在又一实现中,还可以在量化器块420中的量化之后的设置成零块422中执行设置成零操作。在该实现中,不会出现块410和418。一般地,根据具体实现来提供块410、418、422中的至少之一。

[0130] 然后,在块422的输出端处,获得与图3A中示出的频谱对应的经量化的频谱。然后,该经量化的频谱被输入至熵编码器如图2B中的232,其可以是Huffman编码器或按照例如

USAC标准定义的算术编码器。

[0131] 彼此交替地或者并行地设置的设置成零块410、418、422由频谱分析器424来控制。优选地,频谱分析器包括公知的音调检测器的任何实现或者包括操作地用于将频谱分离成要以高分辨率编码的分量以及要以低分辨率编码的分量的任何不同类型的检测器。在频谱分析器中实现的其他这样的算法可以是语音活动检测器、噪声检测器、话音检测器或根据不同频谱部分的频谱信息或关于分辨率要求的相关联的元数据进行判决的任何其他检测器。

[0132] 图5A示出了图1A的时间频谱转换器100的优选实现,例如按照AAC或USAC来实现。时间频谱转换器100包括由瞬态检测器504控制的加窗器(windower) 502。当瞬态检测器504检测到瞬态时,则从长窗至短窗的切换被用信号发送给加窗器。然后,加窗器502针对交叠的块计算加窗的帧,其中每个加窗的帧通常具有 $2N$ 个值如2048个值。然后,在块变换器506内执行变换,并且该块变换器通常额外地提供抽取(decimation)来使得执行组合的抽取/变换以获得具有 N 个值如MDCT频谱值的频谱帧。从而,对于长窗操作,块506的输入端处的帧包括 $2N$ 个值如2048个值并且频谱帧则具有1024个值。那么,然而,当执行8个短块,并且其中每个短块与长窗相比具有 $1/8$ 的加窗时域值并且每个频谱块相比长块具有 $1/8$ 的频谱值时,对短块执行切换。从而,当该抽取与加窗器的50%的交叠操作组合时,频谱是时域音频信号99的临界采样的版本。

[0133] 随后,参照图5B,图5B示出了图1B的频率再生器116和频谱时间转换器118或图2A的块208、212的组合操作的具体实现。在图5B中,考虑特定的重构频带如图3A的比例因子频带6。该重构频带中的第一频谱部分即图3a的第一频谱部分306被输入至帧构建器/调整器块510。此外,比例因子频带6的经重构的第二频谱部分也被输入至帧构建器/调整器510。此外,比例因子频带6的能量信息如图3B的 E_3 也被输入至块510。重构频带中的经重构的第二频谱部分已经使用源范围通过频率铺片填充被生成,则重构频带对应于目标范围。现在,执行帧的能量调整以因此最终获得具有 N 个值的完整的经重构的帧如例如在图2A的组合器208的输出端处获得的。然后,在块512中,执行逆块变换/插值以在块512的输入端处获得例如124个频谱值的248个时域值。然后,在块514中执行合成加窗操作,块514再次由作为编码音频信号中的边信息被发送的长窗/短窗指示来控制。然后,在块516中,执行与先前时间帧的交叠/相加操作。优选地,MDCT应用50%的交叠,使得对于 $2N$ 个值的每个新的时间帧,最终输出 N 个时域值。50%的交叠由于以下事实而被强烈地优选:其提供临界采样以及由于块516中的交叠/相加操作而提供从一个帧至下一个帧的连续交叉。

[0134] 如图3A中的301处所示,例如对于与图3A的比例因子频带6重合的设想的重构频带,可以不仅在IGF起始频率之下而且还在IGF起始频率之上额外地应用噪声填充操作,然后,噪声填充频谱值也可以被输入至帧构建器/调整器510,并且在该块中还可以应用噪声填充频谱值的调整,或者在被输入至帧构建器/调整器510之前,就已经可以使用噪声填充能量来调整噪声填充频谱值。

[0135] 优选地,IGF操作,即,使用来自其他部分的频谱值进行的频谱铺片填充操作可以被应用于完整的频谱。从而,频谱铺片填充操作不仅可以应用于IGF起始频率之上的高频带中,而且还可以应用于低频带中。此外,不使用频谱铺片填充的噪声填充也不仅可以应用于IGF起始频率之下,而且还可以应用于IGF起始频率之上。然而,已经发现,如图3A中所示,当

噪声填充操作限于IGF起始频率之下的频谱范围时并且当频率铺片填充操作限于IGF起始频率之上的频谱范围时,可以获得高品质和高效音频编码。

[0136] 优选地,目标铺片(TT)(具有大于IGF起始频率的频率)受制于全速率编码器的比例因子频带边界。从其获取信息的源铺片(ST),即,低于IGF起始频率的频率的源铺片不受受制于比例因子频带边界。ST的尺寸应当对应于相关联的TT的尺寸。这使用下面的示例来示出。TT[0]具有10个MDCT区元的长度。这精确地对应于两个后续的SCB的长度(如4+6)。然后,要与TT[0]相关的所有可能的ST也具有10个区元的长度。与TT[0]相邻的第二目标铺片TT[1]具有15个区元I的长度(具有7+8的长度的SCB)。然后,其ST具有15个区元的长度而非TT[0]的10个区元的长度。

[0137] 假如出现不能找到用于具有目标铺片的长度的ST的TT的情况(当例如TT的长度大于可用源范围时),则不计算相关性,并且不将源范围多次拷贝至该TT中(一个接一个进行拷贝以使得第二拷贝的最低频率的频率线在频率上紧紧跟随第一拷贝的最高频率的频率线),直到目标铺片TT被完全填满为止。

[0138] 随后,参照图5C,图5C示出了图1B的频率再生器116或图2A的IGF块202的又一优选实施方式。块522是频率铺片生成器,其不仅接收目标频带ID而且另外还接收源频带ID。示例性地,在编码器侧确定了:图3A的比例因子频带3非常好地适合于重构比例因子频带7。从而,源频带ID将是2,目标频带ID将是7。基于该信息,频率铺片再生器522应用复制或谐波铺片填充操作或任何其他铺片填充操作以生成频谱分量523的原始第二部分。频谱分量的原始第二部分具有与第一组第一频谱部分中包括的频率分辨率相同的频谱分辨率。

[0139] 然后,重构频带如图3A的307的第一频谱部分被输入至帧构建器524,并且原始第二部分523也被输入至帧构建器524。然后,调整器526使用由增益因子计算器528计算出的重构频带的增益因子对经重构的帧进行调整。然而,重要地,帧中的第一频谱部分不受调整器526影响,而是重构帧的仅原始第二部分受调整器526的影响。为此,增益因子计算器528分析源频带或原始第二部分523并且额外地分析重构频带中的第一频谱部分以最终找到正确的增益因子527,使得当设想比例因子频带7时由调整器526输出的经调整的帧的能量具有能量 E_4 。

[0140] 在该背景下,非常重要地,相比于HE-AAC来评估本发明的高频重构准确度。这将相对于图3A中的比例因子频带7来说明。假定现有技术编码器如图13A中示出的编码器将要以高分辨率编码的频谱部分307检测为“丢失的谐波”。然后,将该频谱分量的能量连同关于该重构频带如比例因子频带7的频谱包络信息一起发送至解码器。然后,解码器将重构丢失的谐波。然而,在其处由图13B的现有技术解码器重构丢失的谐波307的频谱值将位于由重构频率390表示的频率处的频带7的中间。从而,本发明避免会由图13D的现有技术解码器引入的频率误差391。

[0141] 在一种实现中,频谱分析器还被实现成:计算第一频谱部分与第二频谱部分之间的相似度并且基于所计算出的相似度针对重构范围内的第二频谱部分确定尽可能地与第二频谱部分匹配的第一频谱部分。然后,在该可变的源范围/目的范围实现中,参数化编码器会额外地将针对每个目的范围指示匹配源范围的匹配信息引入至第二编码表示中。然后,在解码器侧,该信息将由图5C的频谱铺片生成器522使用,图5C示出了基于源频带ID和目标频带ID来生成原始第二部分523。

[0142] 此外,如图3A中所示,频谱分析器被配置成分析直至最大分析频率的频谱表示,最大分析频率仅少许低于采样频率的二分之一,并且优选地是采样频率的至少四分之一或通常更高。

[0143] 如所示出的,编码器在无下采样的情况下进行操作,并且解码器在无上采样的情况下进行操作。换言之,频谱域音频编码器被配置成生成具有奈奎斯特频率的频谱表示,该奈奎斯特频率由与原始输入的音频信号的采样速率来限定。

[0144] 此外,如图3A中所示,频谱分析器被配置成分析频谱表示,该频谱表示始于间隙填充起始频率并且止于由该频谱表示中包括的最大频率表示的最大频率,其中,从最小频率向上延伸至间隙填充起始频率的频谱部分属于第一组频谱部分,并且其中,具有高于间隙填充频率的频率值的另外的频谱部分如304、305、306、307还被包括在第一组第一频谱部分中。

[0145] 如所概述的,频谱域音频解码器112被配置成使得由第一解码表示中的频谱值表示的最大频率等于具有采样速率的时间表示中包括的最大频率,其中,第一组第一频谱部分中的最大频率的频谱值为零或不为零。不管怎样,对于第一组频谱分量中的该最大频率,存在关于比例因子频带的比例因子,不论如在图3A和图3B的背景下所讨论的该比例因子频带中的所有频谱值被设置成零还是不被设置成零,均生成和发送该比例因子。

[0146] 因此,本发明的优点在于:相对于增大压缩效率的其他参数化技术,例如,噪声替代和噪声填充(这些技术专门用于类噪声局部信号内容的有效表示),本发明使得能够对音调分量进行准确的频率再现。目前,没有现有技术通过频谱间隙填充来解决任意信号内容的有效参数化表示,而不限低频带(LF)和高频带(HF)中的固定的先验划分。

[0147] 本发明系统的实施方式改进现有技术方法,从而甚至在低比特率的情况下仍提供高压缩效率、没有或仅有小量的感知烦恼以及全音频带宽。

[0148] 一般系统包括:

- [0149] • 全频带核心编码
- [0150] • 智能间隙填充(铺片填充或噪声填充)
- [0151] • 由音调掩蔽选择的核心中的稀疏音调部分
- [0152] • 全频带的联合立体声对编码,包括铺片填充
- [0153] • 对铺片的TNS
- [0154] • IGF范围内的频谱白化

[0155] 迈向更高效系统的第一步是去除将频谱数据变换成与核心编码的变换域不同的第二变换域的需要。当大多数音频编解码器例如AAC使用MDCT作为基本变换,在MDCT域也执行BWE是有用的。BWE系统的第二要求将是保存音调网格的需要,由此甚至HF音调分量被保存,从而编码音频的品质好于现有系统。为了小心处理上面提到的关于BWE方案的两个要求,提出了称作智能间隙填充(IGF)的新系统。图2B示出了编码器侧的所提出的系统的框图,图2A示出了解码器侧的系统。

[0156] 图6A示出了本发明的另一实现中的用于对编码音频信号进行解码的设备。用于解码的设备包括:频谱域音频解码器602,其用于生成第一组频谱部分的第一解码表示;以及频率再生器604,其连接在频谱域音频解码器602的下游,用于使用第一组第一频谱部分中的第一频谱部分来生成经重构的第二频谱部分。如603处所示,第一频谱部分和第二频谱部

分中的频谱值是频谱预测残差值。为了将这些频谱预测残差值变换成全频谱表示,提供了频谱预测滤波器606。该逆预测滤波器被配置成使用第一组第一频率和经重构的第二频谱部分的频谱残差值来执行关于频率的逆预测。频谱逆预测滤波器606由被包括在编码音频信号中的滤波器信息来配置。图6B示出了图6A的实施方式的更详细的实现。频谱预测残差值603被输入至频率铺片再生器612,该频率铺片再生器612生成关于重构频带或某个第二频率部分的原始频谱值,从而现在具有与第一频谱表示的高分辨率相同的分辨率的该原始数据被输入至频谱整形器614。频谱整形器现在使用比特流中传输的包络信息来对频谱进行整形,然后频谱整形的数据被应用于频谱预测滤波器616,该频谱预测滤波器616使用从编码器经由比特流发送至解码器的滤波器信息607来最终生成全频谱值的帧。

[0157] 在图6B中假定:在编码器侧,在计算包络信息之后,执行经由比特流发送且经由线607使用的滤波器信息的计算。因此,换言之,与图6B的解码器匹配的编码器将首先计算频谱残差值,然后会使用频谱残差值来计算包络信息,例如如图7A中所示。然而,对于某些实现而言,其他实现也是有用的,其中,在在编码器侧执行TNS或TTS滤波之前,计算包络信息。然后,在在块624中执行频谱整形之前,应用频谱预测滤波器622。从而,换言之,在应用频谱整形操作624之前,生成(全)频谱值。

[0158] 优选地,计算复数值TNS滤波器或TTS滤波器。这在图7A中被示出。原始音频信号被输入至复数MDCT块702。然后,在复数域执行TTS滤波器计算和TTS滤波。然后,在块706中,计算IGF边信息,以及还计算任何其他操作如用于编码的频谱分析等。然后,由块706生成的第一组第一频谱部分由708处示出的心理声学模型驱动编码器编码以获得在图7A中的X(k)处表示的第一组第一频谱部分,然后所有这些数据被转发至比特流复用器710。

[0159] 在解码器侧,编码数据被输入至解复用器720来一方面分离IGF边信息,另一方面分离TTS边信息以及第一组第一频谱部分的编码表示。

[0160] 然后,块724用于根据一个或更多个实数值频谱来计算复数频谱。然后,实数值频谱和复数频谱两者均被输入至块726以生成对于重构频带的第二组第二频谱部分中的经重构的频率值。然后,对完整地获得并且铺片填充的全频带帧执行逆TTS操作728,以及在解码器侧,在块730中执行最终的逆复数MDCT操作。从而,复数TNS滤波器信息的使用使得当不仅被应用在核心频带或单独的铺片频带内而且被应用在核心/铺片边界或铺片/铺片边界上时能够自动生成铺片边界处理,最终其重新引入铺片之间的频谱相关性。铺片边界上的该频谱相关性不能仅通过生成频率铺片和对频率铺片的原始数据执行频谱包络调整来获得。

[0161] 图7C示出了原始信号(左侧面板)与不使用TTS的扩展信号的比较。可以看到,存在由750处示出的上频率范围中的扩大部分示出的强伪声。然而,当750处的同一频谱部分与图7C的伪声相关的分量750相比时,在图7E中却不发生这种情况。

[0162] 实施方式或本发明的音频编码系统使用可用比特率的主要份额以在编码器中对信号的仅感知上最相关的结构进行波形编码,并且结果的频谱间隙在解码器中被填充有大致近似于原始频谱的信号内容。非常有限的比特预算被消耗以通过从编码器发送至解码器的专用边信息来控制参数驱动的所谓的频谱智能间隙填充(IGF)。

[0163] 音频信号的存储或传输通常受到严格的比特率约束。在过去,仅当很低的比特率可用时,编码器才被迫大幅度地减小传输音频带宽。现代音频编解码器如今能够通过使用类似于频谱带宽复制(SBR) [1]的带宽扩展(BWE)方法对宽带信号进行编码。这些算法依赖

于高频内容 (HF) 的参数化表示以及参数驱动的后处理的应用, 高频内容 (HF) 根据解码信号的波形编码低频部分 (LF) 借助于至 HF 频谱区域的移动 (“修补”) 来生成。在 BWE 方案中, 高于给定的所谓的交叉频率的 HF 频率区域的重构通常基于频谱修补。通常, HF 区域包括多个相邻的补片, 并且这些补片中每一个源自低于给定的交叉频率的 LF 频谱的带通 (BP) 区域。现有技术系统通过将相邻子带系数组从源区域复制到目标区域来有效地执行滤波器组表示内的修补。

[0164] 如果在滤波器组或时间频率变换域中实现 BWE 系统, 则仅存在一个有限的可能性来控制带宽扩展信号的时间形状。通常, 时间粒度受限于在相邻变换窗之间使用的跳距。这可以导致 BWE 频谱范围内的不想要的预回声和后回声。

[0165] 根据感知音频编码, 已知的是, 音频信号的时间包络的形状可以通过使用频谱滤波技术如时间包络整形 (TNS) [14] 来恢复。然而, 根据现有技术所知的 TNS 滤波器是关于实数值频谱的实数值滤波器。关于实数值频谱的这样的实数值滤波器可以由混叠的伪声严重地影响, 尤其在基本的实数变换是修正离散余弦变换 (MDCT) 的情况下。

[0166] 时间包络铺片整形对复数值频谱应用复数滤波, 比如通过例如复数修正离散余弦变换 (CMDCT) 得到的。从而, 避免了混叠的伪声。

[0167] 时间铺片整形包括:

[0168] • 复数滤波器系数估计以及在编码器处对原始信号频谱应用平坦化滤波器

[0169] • 边信息中的滤波器系数的传输

[0170] • 在解码器处对铺片填充的经重构的频谱应用整形滤波器

[0171] 本发明通过沿着频率方向的线性预测来扩展从音频变换编码尤其时间噪声整形 (TNS) 已知的现有技术, 以在带宽扩展的背景下以修正的方式来使用。

[0172] 此外, 本发明的带宽扩展算法基于智能间隙填充 (IGF), 但是与依赖于信号的实数值临界采样的 MDCT 表示的 IGF 标准配置相反, 利用过采样的复数值变换 (CMDCT)。CMDCT 可以被看作每个复数值频谱系数的实部的 MDCT 系数与虚部的 MDST 系数的组合。

[0173] 虽然在 IGF 的背景下描述的新的方法, 但是本发明的处理可以结合基于音频信号的滤波器组表示的任何 BWE 方法来使用。

[0174] 在该新颖的背景下, 沿着频率方向的线性预测不被用作时间噪声整形, 而是被用作时间铺片整形 (TTS) 技术。该重命名通过以下事实来证明: 与通过现有技术的感知变换编解码器中的 TNS 进行量化噪声整形相反, 铺片填充的信号分量通过 TTS 来进行时间整形。

[0175] 图 7A 示出了使用 IGF 和新的 TTS 方法的 BWE 编码器的框图。

[0176] 所以基本的编码方案工作如下:

[0177] - 计算时域信号 $x(n)$ 的 CMDCT 以获得频域信号 $X(k)$

[0178] - 计算复数值 TTS 滤波器

[0179] - 获得用于 BWE 的边信息并且去除必须由解码器复制的频谱信息

[0180] - 使用心理声学模块 (PAM) 应用量化

[0181] - 存储/发送数据, 仅实数值 MDCT 系数被发送

[0182] 图 7B 示出了相应的解码器。其主要逆向进行在编码器中执行的步骤。

[0183] 在此, 基本的解码方案工作如下:

[0184] - 根据 MDCT 值估计 MDST 系数 (该处理增加一个块解码器延迟), 并且将 MDCT 系数和

MDST系数组合成复数值CMDCT系数

[0185] -使用其后处理执行铺片填充

[0186] -使用所传输的TTS滤波器系数应用逆TTS滤波

[0187] -计算逆CMDCT

[0188] 注意,可替代地,如果在编码器一致地颠倒TTS分析和IDG参数估计,则在解码器中还可以颠倒TTS合成和IGF后处理的顺序。

[0189] 对于高效变换编码,优选地,必须使用所谓的近似20ms的“长块”来实现合理的变换增益。如果这样的长块内的信号包括瞬态,则由于铺片填充在经重构的频谱频带中出现听得见的预回声和后回声。图7C示出了由于IGF损害瞬态的典型的预回声和后回声效果。在图7C的左侧面板上,示出了原始信号的频谱图,而在右侧面板上,示出了不具有本发明的TTS滤波的铺片填充的信号的频谱图。在该示例中,核心频带与铺片填充的频带之间的IGF起始频率 $f_{IGFstart}$ 或 f_{Split} 被选择为 $f_s/4$ 。在图7C的右侧面板中,围绕瞬态可见不同的预回声和后回声,在复制的频率区域的上频谱结束处尤为突出。

[0190] TTS模块的主要任务是限制瞬态附近的这些不想要的信号分量,从而将它们隐藏在由人类感知的时间掩蔽效果支配的时间区域中。因此,所需要的TTS预测系数被计算并且使用“前向预测”被应用于CMDCT域中。

[0191] 在将TTS和IGF组合到编解码器中的实施方式中,重要的是,对某些TTS参数和IGF参数进行排列以使得IGF铺片完全地被或者不被一个TTS滤波器(平坦或整形滤波器)滤波。因此,所有TTSstart[...]或TTSstop[...]频率不应当被包括在IGF铺片内,而是被排列至相应的 $F_{IGF}...$ 频率。图7D示出了一组三个TTS滤波器的TTS和IGF操作区域的示例。

[0192] TTS终止频率被调整成IGF工具的终止频率,其高于 $f_{IGFstart}$ 。如果TTS使用多于一个滤波器,则必须确保:两个TTS滤波器之间的交叉频率必须匹配IGF分割频率。否则,一个TTS子滤波器将超出 $f_{IGFstart}$,导致不想要的伪声例如过整形。

[0193] 在图7A和图7B中描绘的实现变型中,必须额外注意,在解码器IGF中正确地调整能量。在以下情况尤其如此:在TTS和IGF处理的过程中具有不同的预测增益的不同的TTS滤波器被应用于一个IGF铺片的源区域(如平坦化滤波器)和目标频谱区域(如整形滤波器,其不是所述平坦化滤波器的精确副本)。在这种情况下,两个所应用的TTS滤波器的预测增益比率不再等于1,并且因此必需应用以该比率的能量调整。

[0194] 在替选的实现变型中,IGF后处理和TTS的顺序被颠倒。在解码器中,这表示通过IGF后处理的能量调整在TTS滤波之后被计算,并且因此是合成变换之前的最终处理步骤。因此,不管在编码期间不同的TTS滤波器增益被应用于一个铺片,通常总是通过IGF处理正确地调整最终能量。

[0195] 在解码器侧,TTS滤波器系数再次被应用在全频谱即通过再生频谱扩展的核心频谱上。TTS的应用对于以下是必需的:形成再生频谱的时间包络以再次匹配原始信号的包络。所以,示出的预回声被减少。另外,如同传统的TNS,其仍然在时间上对信号中的低于 $f_{IGFstart}$ 的量化噪声进行整形。

[0196] 在传统的编码器中,对音频信号的频谱修补(例如,SBR)会破坏补片边界处的频谱相关性,并且因此通过引入扩散来损害音频信号的时间包络。因此,对残差信号执行IGF铺片填充的另一益处在于:在应用TTS整形滤波器之后,铺片边界无缝地相关,导致对信号的

更逼真的时间再现。

[0197] 在图7E中示出了相应处理的信号的结果。与未滤波的版本(图7C,右侧面板)相比,TTS滤波的信号显示出了很好地减少了不想要的预回声和后回声(图7E,右侧面板)。

[0198] 此外,如所讨论的,图7A示出了与图7B的解码器或图6A的解码器匹配的编码器。基本上,用于对音频信号进行编码的设备包括时间频谱转换器如702,其用于将音频信号转换成频谱表示。频谱表示可以是实数值频谱表示,或者如块702中所示的复数值频谱表示。此外,用于执行关于频率的预测的预测滤波器如704被设置为生成频谱残差值,其中,预测滤波器704由从音频信号得到并且被转发至比特流复用器710的预测滤波器信息来限定,如图7A中的714处所示。此外,还设置有音频编码器如心理声学驱动的音频编码器704。音频编码器被配置成用于对频谱残差值的第一组第一频谱部分进行编码以获得编码的第一组第一频谱值。另外,参数化编码器如图7A中的706处所示的参数化编码器被设置成用于对第二组第二频谱部分进行编码。优选地,与第二组第二频谱部分相比,第一组第一频谱部分以较高的频谱分辨率被编码。

[0199] 最后,如图7A中所示,输出接口被设置成用于将包括参数化地编码的第二组第二频谱部分、编码的第一组第一频谱部分和如图7A中的714处示为“TTS边信息”的滤波器信息的编码信号输出。

[0200] 优选地,预测滤波器704包括滤波器信息计算器,其被配置成使用频谱表示的频谱值来计算滤波器信息。此外,预测滤波器被配置成使用用于计算滤波器信息的频谱表示的相同的频谱值来计算频谱残差值。

[0201] 优选地,TTS滤波器704以与根据AAC标准应用TNS工具的已知的现有技术音频编码器相同的方式被配置。

[0202] 随后,在图8A至图8E的背景下来讨论使用双通道解码的另外的实现。此外,参照图2A、图2B的背景下的相应的元素的描述(联合通道编码228和联合通道解码204)。

[0203] 图8A示出了用于生成解码的双通道信号的音频解码器。音频解码器包括用于对编码的双通道信号进行解码以获得第一组第一频谱部分的4个音频解码器802,并且还包括用于提供第二组第二频谱部分的参数化数据以及额外地用于标识第二频谱部分的第一双通道表示或不同的第二双通道表示的双通道标识的参数化解码器804。另外,频率再生器806被设置成用于根据第一组第一频谱部分的第一频谱部分以及第二部分的参数化数据和第二部分的双通道标识来再生第二频谱部分。图8B示出了源范围和目的范围中的双通道表示的不同组合。源范围可以按照在第一双通道表示,并且目的范围也可以按照第一双通道表示。可替代地,源范围可以按照第一双通道表示,而目的范围可以按照第二双通道表示。此外,如图8B的第三列中所示,源范围可以按照第二双通道表示中并且目的范围可以按照第一双通道表示。最后,源范围和目的范围两者可以按照第二双通道表示。在一种实施方式中,第一双通道表示是其中双通道信号的两个通道被单独地表示的分离的双通道表示。然后,第二双通道表示是其中双通道表示的两个通道被联合地表示,即,按照输出至对应扬声器的需求而需要进一步的处理或表示变换来重新计算分离的双通道表示的联合表示。

[0204] 在一种实现中,第一双通道表示可以是左/右(L/R)表示,而第二双通道表示是联合立体声表示。然而,除了左/右或M/S或立体声预测以外的其他双通道表示也可以被应用并且用于本发明。

[0205] 图8C示出了通过图8A的音频解码器执行的操作的流程图。在步骤812中,音频解码器802执行源范围的解码。相对于图3A,源范围可以包括比例因子频带SCB1至SCB3。此外,可以存在每个比例因子频带的双通道标识,并且比例因子频带1可以例如按照第一表示(如L/R),而第三比例因子频带可以按照第二双通道表示如M/S或预测下混合/残差。从而,步骤812可以产生不同频带的不同表示。然后,在步骤814中,频率再生器806被配置成用于选择用于频率再生的源范围。在步骤816中,频率再生器806然后检查源范围的表示,并且在块818中,频率再生器806将源范围的双通道表示与目标范围的双通道表示相比较。如果两个表示相同,则频率再生器806针对双通道信号的每个通道提供单独的频率再生。然而,当在块818中检测的两个表示不同时,则采用信号流824,并且块822根据源范围计算其他双通道表示,并且将该计算出的其他双通道表示用于再生目标范围。从而,图8A的解码器使得可以使用按照第一双通道表示的源范围来再生被表示为具有第二双通道标识的目的范围。当然,本发明还使得能够使用具有相同的双通道标识的源范围来再生目标范围。并且另外地,本发明使得能够:再生具有表示联合双通道表示的双通道标识的目标范围,并且然后将该表示变换成用于存储或传输至双通道信号的对应的扬声器所需的分离的通道表示。

[0206] 强调的是,双通道表示的两个通道可以是两个立体声通道如左通道和右通道。然而,该信号还可以是具有例如五个通道和超低音扬声器通道或具有甚至更多通道的多通道信号。然后,可以执行在图8A至图8E的背景下讨论的成对双通道处理,其中所述对可以是例如左通道和右通道、左环绕通道和右环绕通道以及中心通道和LFE(超低音扬声器)通道。为了通过三个双通道处理过程表示例如六个输入通道,可以使用任何其他配对。

[0207] 图8D示出了与图8A对应的本发明的解码器的框图。源范围或核心解码器830可以对应于音频解码器802。其他块832、834、836、838、840、842、846可以是图8A的频率再生器806的部分。具体地,块832是表示变换器,其用于对各个频带中的源范围表示进行变换,使得在块832的输出处一方面存在按照第一表示的源范围的完整组以及另一方面存在按照第二双通道表示的源范围的完整组。这两个完整的源范围表示可以被存储在用于源范围的两个表示的存储器834中。

[0208] 然后,块836应用频率铺片生成,其使用源范围ID作为输入以及额外地使用目标范围的双通道ID作为输入。基于目标范围的双通道ID,频率铺片生成器访问存储器834并且接收与被输入至835处的频率铺片再生器的目标范围的双通道ID匹配的源范围的双通道表示。从而,当目标范围的双通道ID指示联合立体声处理时,则频率铺片生成器836访问存储器834以获得由源范围ID 833表示的源范围的联合立体声表示。

[0209] 频率铺片生成器836针对每个目标范围执行该操作,并且频率铺片生成器的输出使得呈现由双通道标识所标识的通道表示的每个通道。然后,通过包络调整器838执行包络调整。在由双通道标识所标识的双通道域中执行包络调整。为此,需要包络调整参数,并且这些参数以所描述的相同的双通道表示被从编码器发送至解码器。当要由包络调整器处理的目标范围的双通道标识具有表示与该目标范围的包络数据不同的双通道表示的双通道标识时,则参数变换器840将包络参数变换成所需的双通道表示。当例如一个频带的双通道标识表示联合立体声编码时并且当该目标范围的参数已经作为L/R包络参数被发送时,则参数变换器根据所描述的L/R包络参数计算联合立体声包络参数,使得正确的参数化表示用于目标范围的频谱包络调整。

[0210] 在另一优选实施方式中,当联合立体声用于目标频带时,包络参数已经作为联合立体声参数被发送。

[0211] 当假定到包络调整器838的输入是一组具有不同的双通道表示的目标范围时,则包络调整器838的输出也是按照不同的双通道表示的一组目标范围。当目标范围具有联合表示如M/S时,则该目标范围由表示变换器842处理以计算存储或传输至扬声器所需的分离的表示。然而,当目标范围已经具有分离的表示时,采用信号流844,并且表示变换器842被旁路。在块842的输出处,作为分离的双通道表示的双通道频谱表示被获得,其然后还可以如由块846所示被进一步处理,其中,该进一步的处理可以是例如频率/时间转换或任何其他所需的处理。

[0212] 优选地,第二频谱部分对应于频率频带,并且双通道标识被设置成与图8B的表格对应的旗标数组,其中,每个频带存在一个旗标。然后,参数化解码器被配置成检查是否设置旗标并且根据旗标来控制频率再生器106以使用第一频谱部分的第一表示或第二表示。

[0213] 在一种实施方式中,仅始于图3A的IGF起始频率309的重构范围具有关于不同重构频带的双通道标识。在又一种实施方式中,这也适用于低于IGF起始频率309的频率范围。

[0214] 在又一种实施方式中,可以通过相似度分析来自适应地确定源频带标识和目标频带标识,然而,当存在源范围到目标范围的固定关联时,还可以应用本发明的双通道处理。源范围可以用于关于频率使用类似于从高效AAC处理知道的多个补片的处理的两个或更多个频率铺片填充通过谐波频谱铺片填充操作或复制频率铺片填充操作来重新创建较宽的目标范围。

[0215] 图8E示出了用于对双通道音频信号进行编码的音频编码器。该编码器包括时间频谱转换器860,其用于将双通道音频信号转换成频谱表示。此外,还包括频谱分析器866,其用于将双通道音频通道音频信号转换成频谱表示。此外,频谱分析器866还被设置成用于执行分析以便确定哪些频谱部分要以高分辨率编码,即,找出第一组第一频谱部分以及额外地找出第二组第二频谱部分。

[0216] 此外,双通道分析器864被设置成用于分析第二组第二频谱部分以确定用于标识第一双通道表示或第二双通道表示的双通道标识。

[0217] 取决于双通道分析器的结果,按照第二频谱表示的频带使用第一双通道表示或第二双通道表示被参数化,并且这通过参数编码器868来执行。核心频率范围,即,低于图3A的IGF起始频率309的频带由核心编码器870来编码。块868和870的结果被输入至输出接口872。如所表示的,双通道分析器提供高于IGF起始频率或全部频谱范围的每个频带的双通道标识,并且该双通道标识还被转发至输出接口872以使得该数据还被包括在由输出接口872输出的编码信号873中。

[0218] 此外,优选地,音频编码器包括带状变换器862。基于双通道分析器862的判决,时间频谱转换器862的输出信号被变换成由双通道分析器具体地由双通道ID 835表示的表示。从而,带状变换器862的输出是一组频带,其中,每个频带可以按照第一双通道表示或不同的第二双通道表示。当本发明被应用于全频带时,即,当源范围和重构范围均由带状变换器处理时,频谱分析器860可以分析该表示。然而,可替代地,频谱分析器860还可以分析如由控制线861表示的由时间频谱转换器输出的信号。从而,频谱分析器860可以对带状变换器862的输出或者时间频谱转换器860的输出在由带状变换器862处理之前应用优选的音调

分析。此外,频谱分析器可以对带状变换器862的结果或者时间频谱转换器860的结果应用某个目标范围的最佳匹配的源范围的标识。

[0219] 随后,参照图9A至图9D,其用于示出了在图3A和图3B的背景下已经讨论的能量信息值的优选计算。

[0220] 现有技术音频编码器应用各种技术来使表示给定音频信号的数据量最小化。音频编码器例如USAC[1]应用时间至频率变换如MDCT来获得给定音频信号的频谱表示。这些MDCT系数利用人类听觉系统的心理声学特点被量化。如果降低可用比特率,则量化变得粗糙,引入大量的零频谱值,这在解码器侧处导致听得见的伪声。为了提高感知品质,现有技术解码器使用随机噪声来填充这些零频谱部分。IGF方法从剩余的非零信号获得铺片来填充频谱中的那些间隙。对于解码音频信号的感知品质而言至关重要是:保留频谱包络和频谱系数的能量分布。在此给出的能量调整方法使用发送的边信息来重构音频信号的频谱MDCT包络。

[0221] 在eSBR[15]内,音频信号至少以因子2被下采样,并且频谱的高频部分完全被清零[1,17]。该删除的部分在解码器侧通过参数化技术被替换成eSBR。eSBR暗示使用额外的变换QMF变换,其用于替换空的高频部分并且对音频信号进行重采样[17]。这增加音频编码器的计算复杂度和存储消耗。

[0222] USAC编码器[15]提供使用随机噪声填充频谱洞(零频谱线)的可能性,但是具有以下缺点:随机噪声不能保留瞬态信号的时间精细结构,且也不能保留音调信号的谐波结构。

[0223] eSBR在解码器侧操作的区域完全地被编码器删除[1]。因此,eSBR倾向于删除高频区域中的音调线或使原始信号的谐波结构失真。当eSBR的QMF频率分辨率很低时并且正弦分量的重新插入仅在基本的滤波器组的粗糙分辨率下才可以时,复制的频率范围中的eSBR的音调分量的再生具有很低的精确度。

[0224] eSBR使用调整修补区域的能量的技术,即,频谱包络调整[1]。该技术使用QMF频率时间网格上的发送的能量值来对频谱包络进行重新整形。现有技术不处理部分删除的频谱,并且因为高的时间分辨率,其倾向于需要相对大量的比特来发送适当的能量值或对能量值应用粗糙量化。

[0225] 当其使用如[15]中所描述那样计算的传统的MDCT变换时,IGF的方法不需要额外的变换。

[0226] 在此给出的能量调整方法使用由编码器生成的边信息来重构音频信号的频谱包络。如下面所概述的,该边信息由编码器生成:

[0227] a) 将加窗的MDCT变换应用于输入音频信号[16,章节4.6],可选地计算加窗的MDST或根据所计算出的MDCT来估计加窗的MDST

[0228] b) 对MDCT系数应用TNS/TTS[15,章节7.8]

[0229] c) 计算高于IGF起始频率($f_{IGFstart}$)直到IGF终止频率($f_{IGFstop}$)的每个MDCT比例因子频带的平均能量

[0230] d) 对平均能量值进行量化

[0231] $f_{IGFstart}$ 和 $f_{IGFstop}$ 是用户给定的参数。

[0232] 根据步骤c)和d)所计算出的值被无损编码,并且作为边信息与比特流一起被发送至解码器。

[0233] 解码器接收所发送的值并且使用它们来调整频谱包络。

[0234] a) 对发送的MDCT值进行去量化

[0235] b) 如果被用信号发送,则应用传统的USAC噪声填充

[0236] c) 应用IGF铺片填充

[0237] d) 对所发送的能量值进行去量化

[0238] e) 调整频谱包络比例因子频带状

[0239] f) 如果被用信号发送,应用TNS/TTS

[0240] 设 $\hat{x} \in \mathbb{R}^N$ 为窗长度为 $2N$ 的加窗音频信号的MDCT变换的实数值频谱表示。在[16]中描述了该变换。可选地,编码器对 \hat{x} 可选地应用TNS。

[0241] 在[16,4.6.2]中,描述了比例因子频带中的 \hat{x} 的分解。比例因子频带是一组索引的集合,并且在本文中用scb来表示。

[0242] 每个 scb_k (其中, $k=0,1,2,\dots,\max_sfb$) 的极限由数组 swb_offset [16,4.6.2] 来限定,其中, $swb_offset[k]$ 和 $swb_offset[k+1]-1$ 限定 scb_k 中包括的最低频谱系数和最高频谱系数的第一索引和最后一个索引。将比例因子频带表示为:

[0243] $scb_k := \{swb_offset[k], 1+swb_offset[k], 2+swb_offset[k], \dots, swb_offset[k+1]-1\}$

[0244] 如果编码器使用IGF工具,则用户限定IGF起始频率和IGF终止频率。这两个值被映射到最佳填充比例因子频带索引 $igfStartSfb$ 和 $igfStopSfb$ 。这两者在比特流中被用信号发送给解码器。

[0245] [16]描述了长块变换和短块变换。对于长块,仅一组频谱系数连同比例因子被发送至解码器。对于短块,8个短窗连同8组不同的频谱系数被计算。为了节省比特率,这8个短块窗的比例因子由编码器分组。

[0246] 在IGF的情况下,在此给出的方法使用传统的比例因子频带来对被发送至解码器的频谱值进行分组:

$$[0247] \quad E_k = \sqrt{\frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{x}_i^2}$$

[0248] 其中, $k=igfStartSfb, 1+igfStartSfb, 2+igfStartSfb, \dots, igfEndSfb$ 。

[0249] 对于量化,计算 $\hat{E}_k = nINT(4\log_2(E_k))$ 。所有值 \hat{E}_k 被发送至解码器。

[0250] 我们假定编码器决定对 num_window_group 个比例因子组进行分组。用 w 来表示作为8个短窗的索引的集合 $\{0,1,2,\dots,7\}$ 的分组分解。 w_l 表示 w 的第 l 个子集,其中, l 表示窗组的索引,且 $0 \leq l < num_window_group$ 。

[0251] 对于短块计算,用户定义的IGF起始/终止频率被映射到适当的比例因子频带。然而,为了简单起见,针对短块表示: $k=igfStartSfb, 1+igfStartSfb, 2+igfStartSfb, \dots, igfEndSfb$ 。

[0252] IGF能量计算使用分组信息来对值 $w_{k,l}$ 进行分组:

$$[0253] \quad E_{k,l} := \sqrt{\frac{1}{|w_l|} \sum_{j \in w_l} \frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{x}_{j,i}^2}$$

[0254] 对于量化, 计算 $\hat{E}_{k,l} = nINT(4\log_2(E_{k,l}))$ 。所有值 $\hat{E}_{k,l}$ 被发送至解码器。

[0255] 上面提到的编码公式使用仅实数值MDCT系数 \hat{x} 进行操作。为了获得IGF范围内的更稳定的能量分布, 即, 降低时间幅值波动, 可以使用替选的方法来计算值 \hat{E}_k 。

[0256] 设 $\hat{x}_r \in \mathbb{R}^N$ 为窗长度为2N的加窗音频信号的MDCT变换的实数值频谱表示, 以及 $\hat{x}_i \in \mathbb{R}^N$ 为音频信号的同一部分的实数值MDST变换的频谱表示。可以精确地计算或根据 \hat{x}_r 估计MDST频谱表示 \hat{x}_i 。 $\hat{c} := (\hat{x}_r, \hat{x}_i) \in \mathbb{C}^N$ 表示加窗音频信号的复数频谱表示, 使得 \hat{x}_r 为其实部, \hat{x}_i 为其虚部。可选地, 编码器对 \hat{x}_r 和 \hat{x}_i 应用TNS。

[0257] 现在, IGF范围中的原始信号的能量可以使用下式来量度:

$$[0258] \quad E_{ok} = \frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{c}_i^2$$

[0259] 重构频带的实数值能量和复数值能量, 即, 在重构IGF范围scb_k时应当在解码器侧使用的铺片使用下式来计算:

$$[0260] \quad E_{tk} = \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{c}_i^2, \quad E_{rk} = \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{x}_{r_i}^2$$

[0261] 其中, tr_k是一组索引—依赖scb_k的关联源铺片范围。在上面的两个公式中, 代替索引组scb_k, 可以使用组 $\overline{scb_k}$ (在本文中随后定义) 来重建tr_k以实现更准确的值E_t和E_r。计算

$$[0262] \quad f_k = \frac{E_{ok}}{E_{tk}}$$

[0263] 如果E_{tk}>0, 否则f_k=0。

[0264] 现在利用 $E_k = \sqrt{f_k E_{rk}}$ 计算E_k的更稳定的版本, 由于仅以MDCT值计算E_k因MDCT值不遵守帕塞瓦尔定理的事实而受损, 因此, 他们不反映频谱值的完整能量信息。按照上面来计算 \hat{E}_k 。

[0265] 如较早指出的, 对于短块, 假定编码器决定对num_window_group个比例因子组进行分组。如上所述, w_l表示w的第l个子集, 其中, l表示窗组的索引, 并且0≤l<num_window_group。

[0266] 再次, 可以计算上面概述的用于计算E_{k,l}的更稳定的版本的替选版本。定义 $\hat{c} := (\hat{x}_r, \hat{x}_i) \in \mathbb{C}^N$, $\hat{x}_r \in \mathbb{R}^N$ 是长度为2N的MDCT变换加窗音频信号以及 $\hat{x}_i \in \mathbb{R}^N$ 是长度为2N的MDST变换加窗音频信号, 计算:

$$[0267] \quad E_{ok,l} = \frac{1}{|w_l|} \sum_{i \in w_l} \frac{1}{|scb_k|} \sum_{i \in scb_k} \hat{c}_{i,l}^2$$

[0268] 类似地, 计算,

$$[0269] \quad E_{tk,l} = \frac{1}{|w_l|} \sum_{i \in w_l} \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{c}_{i,l}^2, \quad E_{rk,l} = \frac{1}{|w_l|} \sum_{i \in w_l} \frac{1}{|scb_k|} \sum_{i \in tr_k} \hat{x}_{r,l}^2,$$

[0270] 并且使用因子 $f_{k,l}$ 来处理,

$$[0271] \quad f_{k,l} = \frac{E_{ok,l}}{E_{tk,l}}$$

[0272] 其用于调整先前计算出的 $E_{rk,l}$:

$$[0273] \quad E_{k,l} = \sqrt{f_{k,l} E_{rk,l}}$$

[0274] 按照上面来计算 $\hat{E}_{k,l}$ 。

[0275] 不仅使用从复数重构频带或从MDCT值得到的重构频带的能量而且还使用来自源范围的能量信息的过程提供改进的能量重构。

[0276] 具体地, 参数计算器1006被配置成使用关于重构频带的能量的信息以及额外地使用关于要用于对重构频带进行重构的源范围的能量的信息来计算关于重构频带的能量信息。

[0277] 此外, 参数计算器1006被配置成计算关于原始信号的复数频谱的重构频带的能量信息 (E_{ok}), 以计算关于要用于对重构频带进行重构的原始信号的复数频谱的实数值部分的源范围的另外的能量信息 (E_{rk}), 其中, 参数计算器被配置成使用能量信息 (E_{ok}) 和另外的能量信息 (E_{rk}) 来计算关于重构频带的能量信息。

[0278] 此外, 参数计算器1006被配置成: 用于确定关于原始信号的复数频谱的经重构的比例因子频带的第一能量信息 (E_{ok}); 用于确定关于要用于对经重构的比例因子频带进行重构的原始信号的复数频谱的源范围的第二能量信息 (E_{tk}); 用于确定关于要用于对经重构的比例因子频带进行重构的原始信号的复数频谱的实数部分的源范围的第三能量信息 (E_{rk}); 用于基于第一能量信息、第二能量信息和第三能量信息中的至少两个之间的关系来确定加权信息; 以及用于使用加权信息对第一能量信息和第三能量信息之一进行加权以获得加权的能量信息并且将加权的能量信息用作关于重构频带的能量信息。

[0279] 用于所述计算的示例如下, 但是鉴于上面的一般原理, 对本领域中的技术人员而言, 将出现很多其他示例:

[0280] A)

$$[0281] \quad f_k = E_{ok}/E_{tk};$$

$$[0282] \quad E_k = \sqrt{f_k * E_{rk}};$$

[0283] B)

$$[0284] \quad f_k = E_{tk}/E_{ok};$$

$$[0285] \quad E_k = \sqrt{(1/f_k) * E_{rk}};$$

[0286] C)

[0287] $f_k = E_{rk}/E_{tk}$;

[0288] $E_k = \sqrt{f_k * E_{ok}}$

[0289] D)

[0290] $f_k = E_{tk}/E_{rk}$;

[0291] $E_k = \sqrt{(1/f_k) * E_{ok}}$

[0292] 所有这些示例承认以下事实：尽管在解码器侧仅处理实数MDCT值，但是由于交叠和相加，实际的计算是隐含地使用复数数字进行的时域混叠消除过程。然而，特别地，与具有重构频带920中的频率的第一频谱部分921不同的频率值的重构频带920的另外的频谱部分922、923的铺片能量信息的确定918依赖于实数MDCT值。因此，被发送至解码器的能量信息通常会小于关于原始信号的复数频谱的重构频带的能量信息 E_{ok} 。例如，对于上面的情况C，这表示因子 f_k (加权信息) 会小于1。

[0293] 在解码器侧，如果IGF工具被用信号通知为打开，则所发送的值 \hat{E}_k 根据比特流来获得，并且应当针对所有 $k = \text{igfStartSfb}, 1 + \text{igfStartSfb}, 2 + \text{igfStartSfb}, \dots, \text{igfEndSfb}$ 使用 $E_k = 2^{\frac{1}{4}\hat{E}_k}$ 来去量化。

[0294] 解码器将所发送的MDCT值去量化成 $x \in \mathbb{R}^N$ ，并计算剩余的残留能量：

$$[0295] \quad sE_k := \sum_{i \in scb_k} x_i^2,$$

[0296] 其中， k 在上面定义的范围。

[0297] 我们表示 $\overline{scb_k} = \{i | i \in scb_k \wedge x_i = 0\}$ 。该集合包括已经由编码器量化成零的比例因子频带 scb_k 的所有索引。

[0298] IGF获得子带方法 (在此未描述) 用于通过使用所发送的MDCT的非零值来填充由于编码器侧处的MDCT频谱值的粗糙量化导致的频谱间隙。 x 会还包括替换所有先前清零的值的值。铺片能量由下式来计算：

$$[0299] \quad tE_k := \sum_{i \in \overline{scb_k}} x_i^2,$$

[0300] 其中， k 在上面定义的范围。

[0301] 在重构频带中丢失的能量由下式来计算：

$$[0302] \quad mE_k := |scb_k| E_k^2 - sE_k$$

[0303] 以及用于调整的增益因子由下式来获得：

$$[0304] \quad g := \begin{cases} \sqrt{\frac{mE_k}{tE_k}} & \text{如果 } (mE_k > 0 \wedge tE_k > 0) \\ 0 & \text{其他} \end{cases}$$

[0305] 其中， $g' = \min(g, 10)$ 。

[0306] 使用该增益因子的频谱包络调整是：对于所有 $i \in \overline{scb_k}$ 且 k 在上面定义的范围

内，

[0307] $x_i := g' x_i$ 。

[0308] 这将x的频谱包络重新整形成原始频谱包络 \hat{x} 的形状。

[0309] 对于短窗序列，所上面所概述的所有计算在原理上保持相同，但是考虑比例因子频带的分组。我们将根据比特流获得的去量化的分组的能量值表示为 $E_{k,l}$ 。计算

$$[0310] \quad sE_{k,l} := \frac{1}{|w_l|} \sum_{j \in w_l} \sum_{i \in scb_{j,k}} x_{j,i}^2$$

[0311] 以及

$$[0312] \quad pE_{k,l} := \frac{1}{|w_l|} \sum_{j \in w_l} \sum_{i \in \overline{scb_{j,k}}} x_{j,i}^2$$

[0313] 索引j描述短块序列的窗索引。

[0314] 计算 $mE_{k,l} := |scb_k| E_{k,l}^2 - sE_{k,l}$

$$[0315] \quad \text{以及 } g := \begin{cases} \sqrt{\frac{mE_{k,l}}{pE_{k,l}}} & \text{如果 } (mE_{k,l} > 0 \wedge pE_{k,l} > 0) \\ 0 & \text{其他} \end{cases}$$

[0316] 其中， $g' = \min(g, 10)$

[0317] 对于所有 $i \in \overline{scb_{k,l}}$ 。

[0318] 应用 $x_{j,i} := g' x_{j,i}$ 。

[0319] 对于低比特率应用，值 E_k 的成对分组在不丢失太多精度的情况下是可以的。该方法仅由长块应用：

$$[0320] \quad E_{k \gg 1} = \sqrt{\frac{1}{|scb_k \cup scb_{k+1}|} \sum_{i \in scb_k \cup scb_{k+1}} \hat{x}_i^2}$$

[0321] 其中， $k = \text{igfStartSfb}, 2 + \text{igfStartSfb}, 4 + \text{igfStartSfb}, \dots, \text{igfEndSfb}$

[0322] 再次，在量化之后，所有值 $E_{k \gg 1}$ 被发送至解码器。

[0323] 图9A示出了用于对包括第一组第一频谱部分的编码表示以及表示第二组第二频谱部分的频谱能量的参数化数据的编码表示的编码音频信号进行解码的设备。在图9A中在901a处表示第一组第一频谱部分，在图9A的901b处表示参数化数据的编码表示。提供音频解码器900用于对第一组第一频谱部分的编码表示901a进行解码以获得经解码的第一组第一频谱部分904，并且用于对参数化数据的编码表示进行解码以获得关于表示个体重构频带的个体能量的第二组第二频谱部分的解码的参数化数据902，其中，第二频谱部分位于重构频带中。此外，提供频率再生器906用于对包括第二频谱部分的重构频带的频谱值进行重构。频率再生器906使用第一组第一频谱部分的第一频谱部分和关于重构频带的个体能量信息，其中，重构频带包括第一频谱部分和第二频谱部分。频率再生器906包括计算器912，其用于确定包括具有重构频带中的频率的第一频谱部分的累积能量的残留能量信息。此外，频率再生器906包括计算器918，其用于确定重构频带的另外的频谱部分的铺片能量信

息,而频率值不同于第一频谱部分,其中,这些频率值具有重构频带中的频率,其中,另外的频谱部分要使用与重构频带中的第一频谱部分不同的第一频谱部分通过频率再生来生成。

[0324] 频率再生器906还包括计算器914,其用于确定重构频带中的丢失能量,并且计算器914使用重构频带的个体能量以及由块912生成的残留能量来操作。此外,频率再生器906包括频谱包络调整器916,其用于基于丢失能量信息和由块918生成的铺片能量信息对重构频带中的另外的频谱部分进行调整。

[0325] 参照图9C,图9C示出了某个重构频带920。该重构频带包括在921处示意性地示出的重构频带中的第一频谱部分如图3a中的第一频谱部分306。此外,重构频带920中的其余频谱值要使用例如来自低于图3a的智能间隙填充起始频率的比例因子频带1、2、3的源区域来生成。频率再生器906被配置成用于生成第二频谱部分922和923的原始频谱值。然后,如图9c中所示,计算增益因子 g 以最终对频率频带922、923中的原始频谱值进行调整来获得重构频带920中的经重构和调整的第二频谱部分,该经重构和调整的第二频谱部分现在与第一频谱部分921具有相同的频谱分辨率即相同的线距离。重要的是要理解,图9C中的921处示出的重构频带中的第一频谱部分由音频解码器900来解码,而不受图9b的包络调整执行块916的影响。替代地,921处所表示的重构频带中的第一频谱部分按原样留下,因为第一频谱部分由全带宽或全速率音频解码器900经由线904来输出。

[0326] 随后,对使用实数数字的特定示例进行讨论。由块912计算出的剩余的残留能量为例如5个能量单位,并且该能量是第一频谱部分921中的示例性地所表示的4个频谱线的能量。

[0327] 此外,与图3B或图3A的比例因子频带6对应的重构频带的能量值 E_3 等于10个单位。重要地,能量值不仅包括频谱部分922、923的能量,而且包括在编码器侧即在使用例如通道掩蔽执行频谱分析之前计算出的重构频带920的全部能量。因此,10个能量单元覆盖重构频带中的第一频谱部分和第二频谱部分。则,假定块922、923的源范围数据或块922、923的原始目标范围数据的能量等于8个能量单元。从而,计算出5个单元的丢失能量。

[0328] 基于丢失能量除以铺片能量 tE_k ,计算出增益因子0.79。然后,将第二频谱部分922、923的原始频谱线乘以所计算出的增益因子。从而,仅对第二频谱部分922、923的频谱值进行调整,而第一频谱部分921的频谱线不受该包络调整影响。在乘以第二频谱部分922、923的原始频谱值之后,计算完整重构频带,其包括重构频带中的第一频谱部分,并且包括重构频带920中的第二频谱部分922、923中的频谱线。

[0329] 优选地,用于生成频带922、923中的原始频谱数据的源范围相对于频率而言低于IGF起始频率309,而重构频带920高于IGF起始频率309。

[0330] 此外,优选地,重构频带边界与比例因子频带边界重合。从而,在一种实施方式中,重构频带具有核心音频解码器的相应的比例因子频带的尺寸,或者将其设定尺寸使得:当应用能量配对时,重构频带的能量值提供等于或大于2的整数个比例因子频带的能量。从而,当假定对于比例因子频带4、比例因子频带5和比例因子频带6执行能量累积时,则重构频带920的较低频率边界等于比例因子频带4的较低边界,而重构频带920的较高频率边界与比例因子频带6的较高边界重合。

[0331] 随后,讨论图9D以示出图9A的解码器的另外的功能。音频解码器900接收与第一组频谱部分的第一频谱部分对应的去量化的频谱值,并且另外地,将如图3B中示出的比例因

子频带的比例因子提供至逆缩放块940。逆缩放块940提供低于图3A的IGF起始频率309的所有第一组第一频谱部分以及额外地提供高于IGF起始频率的第一频谱部分,即,都位于图9D中的941处示出的重构频带的图3A的第一频谱部分304、305、306、307。此外,源频带中的用于重构频带中的频率铺片填充的第一频谱部分被提供至包络调整器/计算器942,并且该块额外接收被提供为如图9D中的943处所示的编码音频信号的参数化边信息的关于重构频带的能量信息。然后,包络调整器/计算器942提供图9B和图9C的功能,并且最后输出重构频带中的第二频谱部分的经调整的频谱值。重构频带中的第二频谱部分的这些经调整的频谱值922、923和图9D中的线941表示的重构频带中的第一频谱部分921联合地表示重构频带的完整的频谱表示。

[0332] 随后,参照图10A和图10B,图10A和图10B用于说明用于对音频信号进行编码以提供或生成编码音频信号的音频编码器的优选实施方式。该编码器包括时间/频谱转换器1002,其给频谱分析器1004馈送,并且频谱分析器1004一方面与参数计算器1006连接并且另一方面与音频编码器1008连接。音频编码器1008提供第一组第一频谱部分的编码表示,并且不覆盖第二组第二频谱部分。另一方面,参数计算器1006提供覆盖第一频谱部分和第二频谱部分的重构频带的能量信息。此外,音频编码器1008被配置成生成具有第一频谱分辨率的第一组第一频谱部分的第一编码表示,其中,音频编码器1008提供由块1002生成的频谱表示的所有频带的比例因子。另外,如图3B中所示,编码器提供相对于频率而言位于如图3A所示的IGF起始频率309之上的至少重构频带的能量信息。从而,对于优选地与比例因子频带或与比例因子频带的组重合的重构频带,给出了两个值,即,来自音频编码器1008的相应的比例因子以及额外地由参数计算器1006输出的能量信息。

[0333] 优选地,音频编码器具有具有不同的频率带宽即具有不同数目的频谱值的比例因子频带。因此,参数化计算器包括归一化器1012,其用于关于特定重构频带的带宽对不同带宽的能量进行归一化。为此,归一化器1012接收频带中的能量以及频带中的大量频谱值作为输入,然后归一化器1012输出归一化能量每重构/比例因子频带。

[0334] 此外,图10A的参数化计算器1006a包括能量值计算器,其用于如图10A中的线1007所示从核心或音频编码器1008接收控制信息。该控制信息可以包括关于由音频编码器使用的长/短块的信息和/或分组信息。因此,虽然关于长/短块的信息以及关于短窗的分组信息与“时间”分组有关,但是分组信息可以还指代频谱分组,即,将两个比例因子频带分成单个重构频带。因此,当已经对仅频谱部分分组时,能量值计算器1014输出覆盖第一频谱部分和第二频谱部分的每个分组的频带的单个能量值。

[0335] 图10D示出了用于实现频谱分组的又一实施方式。为此,块1016被配置成用于计算两个相邻频带的能量值。然后,在块1018中,将两个相邻频带的能量值相比较,并且当能量值与例如由阈值限定的相比未有明显不同或有较小程度的不同时,则如块1020中所示,生成两个频带的单个(经归一化的)值。如由线1019所示,块1018可以被旁路。此外,由块1020执行的关于两个或更多个频带的单个值的生成可以由编码器比特率控制器1024来控制。从而,当比特率要降低时,编码比特率控制器1024控制块1020来生成两个或更多个频带的单个归一化的值,即使块1018中的比较未被允许对能量信息值进行分组。

[0336] 在音频编码器执行两个或更多个短窗的分组的情况下,也对于能量信息应用该分组。当核心编码器执行两个或更多个短块的分组时,则对于这些两个或更多个块,仅单组比

例因子被计算和发送。那么,在解码器侧,音频解码器对于两个分组的窗应用同一组比例因子。

[0337] 关于能量信息计算,对于两个或更多个短窗累积重构频带中的频谱值。换言之,这表示短块和随后的短块的某个重构频带中的频谱值被累积在一起,以及对于覆盖两个短块的该重构频带而言,仅单个能量信息被发送。那么,在解码器侧,对于每个短块不单独执行关于图9a至图9d讨论的包络调整,而是对于一组分组的短窗,一起执行关于图9a至图9d讨论的包络调整。

[0338] 然后,再次应用相应的归一化,使得即使已经执行了频率中的任何分组或者时间中的分组,归一化容易地使得能够:对于解码器侧的能量值信息计算,一方面仅能量值信息以及重构频带中或一组分组重构频带中的频谱线的量必须已知。

[0339] 在现有技术BWE方案中,高于给定的所谓的交叉频率的HF频谱区域的重构通常基于频谱修补。通常,HF区域包括多个相邻的补片,并且这些补片中的每一个来源于低于给定的交叉频率的LF频谱的带通(BP)区域。在信号的滤波器组表示内,这样的系统将LF频谱中的一组相邻的子带系数复制到目标区域中。所选择的组的边界通常是系统依赖的,而非信号依赖的。对于一些信号内容,该静态修补选择可以导致经重构的信号的不愉快的音品和着色。

[0340] 其他方法通过信号自适应单边频带(SSB)调制将LF信号转换成HF。与[1]相比,这样的方法具有高的计算复杂度,因为它们以高采样速率对时域样本进行操作。此外,修补可以变得不稳定,尤其对于非音调信号(例如,非语音语音),从而现有技术信号自适应修补会对信号引入损害。

[0341] 本发明的方法被称为智能间隙填充(IGF),并且按照其优选配置,其被应用于基于时间频率转换例如修正离散余弦变换(MDCT)的BWE系统。然而,本发明的教导一般例如在基于正交镜像滤波器组(QMF)的系统内近似地适用。

[0342] 基于MDCT的IGF配置的优点是无缝集成到基于MDCT的音频编码器例如MPEG高级音频编码(AAC)中。对于波形音频编码和BWE共享相同的变换显著降低音频编解码器的整体计算复杂度。

[0343] 此外,本发明提供了关于在现有技术自适应修补方案中发现的固有稳定性问题的解决方案。

[0344] 所提出的系统基于以下观察:对于一些信号,无引导的修补选择会导致音品变化和信号着色。如果信号在频谱源区域(SSR)中是音调而在频谱目标区域(STR)中类似于噪声,则通过音调SSR修补类噪声STR会导致不自然的音品。信号的音品也可以变化,因为信号的音调结构会通过修补过程变得错位或甚至被破坏。

[0345] 提出的IGF系统使用互相关性作为特定SSR与特定STR之间的类似性量度来执行智能铺片选择。两个信号的互相关性提供那些信号的相似度的量度以及此外最大相关性的滞后及其符号。因此,基于相关性的铺片选择的方法还可以用于精确地调整复制的频谱的频谱偏移以变得尽可能地接近原始频谱结构。

[0346] 提出的系统的基本贡献是对适当的相似度量度的选择以及此外对铺片选择处理稳定化的技术。提出的技术提供了瞬时信号适应与同时时间稳定性之间的最佳平衡。对于具有SSR和STR的很小相似度并且因此展现出低的互相关性值的信号而言或者当采用含糊

的相似度量度时,提供时间稳定性尤其重要。在这种情况下,稳定化防止自适应铺片选择的伪随机行为。

[0347] 例如,通常造成现有技术BWE的问题的一类信号通过至任意频谱区域的不同能量集中来表征,如图12A(左侧)所示。虽然存在可用的方法来调整目标区域中的经重构的频谱的频谱包络和音调,但是对于一些信号而言,这些方法不能够如图12A(右侧)中所示很好地保留音色。在图12A中示出的示例中,原始信号的目标区域中的高于所谓的交叉频率 f_{xover} (图12A,左侧)的频谱的量值近似线性地减小。相反,在经重构的频谱(图12A,右侧)中,存在被感知为音品着色伪声的不同组的谷峰。

[0348] 新的方法的一个重要步骤是定义一组铺片,在该组铺片中可以进行后续的基于相似度的选择。首先,源区域与目标区域两者之间的铺片边界必须根据彼此来定义。因此,核心编码器的IGF起始频率 f_{IGFstart} 与最高可用频率 f_{IGFstop} 之间的目标区域被划分成任意整数个 n_{Tar} 的铺片,这些铺片中的每一个具有各自预定的尺寸。然后,对于每个目标铺片 $\text{tar}[\text{idx_tar}]$,生成一组相等尺寸的源铺片 $\text{src}[\text{idx_src}]$ 。通过这,确定IGF系统的基本自由度。源铺片的总数目 n_{Src} 通过源区域的带宽来确定,

[0349] $\text{bw}_{\text{src}} = (f_{\text{IGFstart}} - f_{\text{IGFmin}})$

[0350] 其中, f_{IGFmin} 是铺片选择的最低可用频率,使得源铺片的整数个 n_{Src} 符合 bw_{src} 。源铺片的最小数目是0。

[0351] 为了进一步增加选择和调整的自由度,源铺片可以被定义成以0与1之间的交叠因子彼此交叠,其中,0表示无交叠,1表示100%交叠。100%交叠可以暗指仅一个或没有源铺片可用。

[0352] 图12B示出了一组铺片的铺片边界的示例。在这种情况下,所有目标铺片与每个源铺片相关。在该示例中,源铺片按照50%交叠。

[0353] 对于目标铺片,在向上 xcorr_maxLag 个区元的滞后处使用各种源铺片来计算互相关性。对于给定的目标铺片 idx_tar 和源铺片 idx_src , $\text{xcorr_val}[\text{idx_tar}][\text{idx_src}]$ 给出了铺片之间的绝对互相关性的最大值,然而, $\text{xcorr_lag}[\text{idx_tar}][\text{idx_src}]$ 给出了在其处出现最大值的滞后, $\text{xcorr_sign}[\text{idx_tar}][\text{idx_src}]$ 给出了 $\text{xcorr_lag}[\text{idx_tar}][\text{idx_src}]$ 处的互相关性的符号。

[0354] 参数 xcorr_lag 用于控制源铺片与目标铺片之间的匹配的接近度。该参数导致伪声减小,并且有助于更好地保留信号的音品和音色。

[0355] 在一些场景下,可以发生特定目标铺片的尺寸大于可用的源铺片的尺寸。在这种情况下,每当需要时重复可用的源铺片来完全填充目标铺片。仍然还可以执行大的目标铺片与较小的源铺片之间的互相关性,以便根据互相关性滞后 xcorr_lag 和符号 xcorr_sign 来获得源铺片在目标铺片中的最好位置。

[0356] 原始频谱铺片和原始信号的互相关性可能不是被应用于具有强共振峰结构的音频频谱的最适合的相似度量度。频谱的白化去除粗糙的包络信息,从而强调频谱精细结构,其对于评估铺片相似度而言最有益。对于由IGF处理的区域,白化还有助于在解码器处容易地对STR进行包络整形。因此,可选地,在计算互相关性之前,对铺片和源信号进行白化。

[0357] 在其他配置中,使用预定的过程仅对铺片进行白化。所发送的“白化”旗标向解码器表示应当对IGF内的铺片应用相同预定的白化处理。

[0358] 为了对信号进行白化,首先计算频谱包络估计。然后,将MDCT频谱除以频谱包络。可以基于MDCT频谱、MDCT频谱能量、基于MDCT的复数功率频谱或功率频谱估计来估计频谱包络估计。从现在开始,在其上估计包络的信号将被称为基信号。

[0359] 在基于MDCT的复数功率谱或功率谱估计上被计算为基信号的包络具有以下优点:在音调分量上没有时间波动。

[0360] 如果基信号位于能量域,则MDCT谱必须除以包络的平方根以正确地对信号进行白化。

[0361] 存在计算包络的不同方法:

[0362] • 利用离散余弦变换(DCT)来变换基信号,仅保留较低DCT系数(将最高的设置为零),然后计算逆DCT

[0363] • 计算在时域音频帧上计算出的一组线性预测系数(LPC)的频谱包络

[0364] • 利用低通滤波器对基信号进行滤波

[0365] 优选地,选择最后一种方法。对于要求低计算复杂度的应用而言,可以对MDCT频谱的白化进行一些简化:首先,借助于移动平均来计算包络。这每MDCT区元仅需要两个处理器周期。然后,为了避免除法和平方根的计算,谱包络近似于 2^n ,其中, n 是包络的整数对数。在该域中,平方根操作简单地变成移位操作,并且此外可以通过另一移位操作来执行除以包络的除法运算。

[0366] 在计算每个源铺片与每个目标铺片的相关性之后,对于所有 n_{Tar} 个目标铺片,选择具有最高相关性的源铺片来替换它。为了最佳匹配原始谱结构,使用相关性的滞后来将所复制的频谱调制整数个变换区元。在奇数滞后的情况下,额外地通过乘以交替的时间序列 $-1/1$ 来调制该铺片以补偿MDCT内的每个其他频带的频率颠倒表示。

[0367] 图12C示出了源铺片与目标铺片之间的相关性的示例。在该示例中,相关性的滞后为5,所以源铺片必须在BWE算法的拷贝段中朝向较高区元被调制5个区元。另外,当最大相关性值为负时,铺片的符号必须被翻转,以及如上所述的额外的调制导致奇数滞后。

[0368] 所以,被从编码器发送至解码器的全部数目的边信息可以包括以下数据:

[0369] • $tileNum[n_{Tar}]$:每目标铺片的所选择的源铺片的索引

[0370] • $tileSign[n_{Tar}]$:目标铺片的符号

[0371] • $tileMod[n_{Tar}]$:每目标铺片的相关性的滞后

[0372] 铺片修剪和稳定化是IGF中的重要步骤。使用示例来说明其需要和优点,假定静态音调音频信号类似例如稳定的律管音符。如果对于给定的目标区域源铺片总是选自横跨多个帧的同一源区域,则逻辑指出引入最少的伪声。即使假定信号为静态,但是由于其他同样相似的源区域的相似度量度(例如,相关性)会支配相似度结果(例如,互相关性),所以该条件在每个帧中无法很好地保持。这导致相邻帧之间的 $tileNum[n_{Tar}]$ 在两个或三个很相似的选择之间摇摆。这可以是讨厌的音乐噪声如伪声的源。

[0373] 为了消除该类型的伪声,一组源铺片应被修剪,使得源组的剩余成员最大程度地不同。这针对一组源铺片 $S = \{s_1, s_2, \dots, s_n\}$ 实现如下。对于任何源铺片 s_i ,将其与所有其他源铺片相关,找出 s_i 与 s_j 之间的最佳相关性,并且将其存储在矩阵 S_x 。在此, $S_x[i][j]$ 包括 s_i 与 s_j 之间的最大绝对互相关性值。沿着列相加矩阵 S_x ,给出源铺片 s_i 与所有其他源铺片 T 的互相关性的和。

[0374] $T[i] = S_x[i][1] + S_x[i][2] \dots + S_x[i][n]$

[0375] 在此, T 表示源如何类似于其他源铺片的度量。如果对于任何源铺片 i , $T >$ 阈值, 则源铺片 i 可以从一组潜在的源中除去, 因为其与源高度相关。该组铺片中的具有最小相关性的满足等式1中的条件的铺片被选择为该子组的代表性铺片。以这种方式, 确保源铺片彼此最大程度地不同。

[0376] 铺片修剪方法还包括前一帧中使用的经修剪的铺片组的存储器。如果存在用于修剪的备选候选铺片, 则在前一帧中活动的铺片也被保留在下一个帧中。

[0377] 设在帧 K 中的铺片 $\{s_1, s_2, \dots, s_5\}$ 中铺片 S_3 、 S_4 和 S_5 有效, 则在帧 $k+1$ 中, 即使铺片 S_1 、 S_3 和 S_2 竞争为使用与其他铺片最大程度地相关的 S_3 来修剪, 但是 S_3 被保留, 因为它是前一帧中的有用的源铺片, 并且因此将其保留在该组源铺片中对于加强铺片选择中的时间连续性有益。如果被表示为 $T_x[i][j]$ 的源 i 与目标 j 之间的互相关性高, 则优选地应用该方法。

[0378] 用于铺片稳定化的另外方法是: 如果当前帧 k 中的源铺片都不与目标铺片很好地相关, 则保留来自前一帧 $k-1$ 的铺片顺序。如果被表示为 $T_x[i][j]$ 的源 i 与目标 j 之间的互相关性对于所有 i 、 j 而言很低, 则这可能发生。

[0379] 例如, 如果 $T_x[i][j] < 0.6$

[0380] 现在使用的试验性阈值, 则对于该帧 k 的所有 $nTar$ 而言,

[0381] $tileNum[nTar]_k = tileNum[nTar]_{k-1}$ 。

[0382] 上述两种技术极大地减少由于快速改变横跨多个帧之间的组铺片数目而出现的伪声。该铺片修剪和稳定化的另一附加的优点在于: 没有额外的信息需要被发送至解码器, 也不需要改变解码器架构。所提出的铺片修剪是减少频谱区域中的潜在的音乐噪声如伪声或过度噪声的上等方式。

[0383] 图11A示出了用于对编码音频信号进行解码的音频解码器。该音频解码器包括音频(核心)解码器1102, 其用于生成第一组第一频谱部分的第一解码表示, 该解码表示具有第一频谱分辨率。

[0384] 此外, 音频解码器包括参数化解码器1104, 其用于生成第二组第二频谱部分的第二解码表示, 该第二解码表示具有第二频谱分辨率, 第二频谱分辨率小于第一频谱分辨率。此外, 提供频率再生器1106, 其用于接收解码的第一频谱部分作为第一输入1101以及接收参数化信息作为1103处的第二输入, 该参数化信息对于每个目标频率或目标重构频带而言包括源范围信息。然后, 频率再生器1106通过使用由匹配信息标识的来自源范围的频谱值来施加频率再生, 以生成目标范围的频谱数据。然后, 第一频谱部分1101和频率再生器1107的输出均被输入至频谱时间转换器1108, 以最终生成解码的音频信号。

[0385] 虽然音频解码器还可以被实现为任何其他音频解码器如时域或参数化音频解码器, 但是优选地, 音频解码器1102是频谱域音频解码器。

[0386] 如在图11B处所表示的, 频率再生器1106可以包括以下功能: 示出了用于奇数滞后的源范围选择器铺片调制器的块1120、当白化旗标1123被提供时的白化滤波器1122, 并且额外地包括在块1128中示出的使用由块1120或块1122或两个块的合作生成的原始频谱数据实现的频谱包络调整功能。无论如何, 频率再生器1106可以包括对所接收的白化旗标1123有反应的开关1124。当白化旗标被设置时, 用于奇数滞后的源范围选择器/铺片调制器的输出被输入至白化滤波器1122。然而, 然后, 对于某个重构频带不设置白化旗标1123, 则

旁路线1126被激活,使得块1120的输出在不进行任何白化的情况下被提供给频谱包络调整块1128。

[0387] 可存在在比特流中用信号通知的多于一个水平的白化(1123),并且这些水平可以每铺片用信号通知。在存在每铺片用信号通知的三个水平的情况下,它们应当以下面的方式被编码:

```
        bit = readBit(1);
        if(bit == 1) {
            for(tile_index = 0..nT)

                /*与上一帧相同的水平*/

                whitening_level[tile_index] = whitening_level_prev_frame[tile_index];
        } else {

            /*第一铺片: */

[0388]     tile_index = 0;
            bit = readBit(1);
            if(bit == 1) {
                whitening_level[tile_index] = MID_WHITENING;
            } else {
                bit = readBit(1);
                if(bit == 1) {
                    whitening_level[tile_index] = STRONG_WHITENING;
                } else {

                    whitening_level[tile_index] = OFF; /*没有白化*/
```

```
    }  
  }  
  
  /*剩余铺片: */  
  
  bit = readBit(1);  
  if(bit == 1) {  
  
    /*剩余铺片的与第一铺片相同的平坦化水平: */  
  
    /*无需读取的另外的比特: */  
  
    for(tile_index = 1..nT)  
      whitening_level[tile_index] = whitening_level[0];  
  } else {  
  
    /*读取就第一铺片而言的其余铺片的比特: */
```

```
[0389]   for(tile_index = 1..nT) {  
        bit = readBit(1);  
        if(bit == 1) {  
            whitening_level[tile_index] = MID_WHITENING;  
        } else {  
            bit = readBit(1);  
            if(bit == 1) {  
                whitening_level[tile_index] = STRONG_WHITENING;  
            } else {  
  
                whitening_level[tile_index] = OFF; /*没有白化*/  
  
            }  
        }  
    }  
}
```

[0390] MID_WHITENING和STRONG_WHITENING指代不同的白化滤波器(1122),白化滤波器(1122)可以在计算包络(如前面所述)的方式上不同。

[0391] 当仅应用粗糙的频谱铺片选择方案时,解码器侧频率再生器可以由源范围ID 1121控制。然而,当应用精细调整的频谱铺片选择方案时,则额外地提供源范围滞后1119。此外,假设相关性计算提供负的结果,于是,另外,相关性的符号可应用于块1120,使得页面数据频谱线每个都乘以“-1”以计及负符号。

[0392] 从而,如图11A、图11B所讨论的,本发明确信:由于某个目的或目标范围的最佳匹配的源范围在编码器侧被计算并且在解码器侧被应用的事实,所以获得最佳音频品质。

[0393] 图11C是用于对音频信号进行编码某种音频编码器,其包括时间频谱转换器1130、随后连接的频谱分析器1132以及另外地参数计算器1134和核心编码器1136。核心编码器1136输出编码的源范围,以及参数计算器1134输出目标范围的匹配信息。

[0394] 编码的源范围连同关于目标范围的匹配信息一起被发送至解码器,使得图11A中示出的解码器能够执行频率再生。

[0395] 参数计算器1134被配置成用于计算第一频谱部分与第二频谱部分之间的相似度并且用于基于所计算出的相似度对于第二频谱部分确定与第二频谱部分匹配的匹配第一频谱部分。优选地,如图12A、图12B中所示不同的源范围和目标范围的匹配结果确定所选择的匹配对包括第二频谱部分,且参数计算器被配置成用于将识别匹配对的该匹配信息提供到编码音频信号中。优选地,该参数计算器1134被配置成使用第二组第二频谱部分中的预定的目标区域或第一组第一频谱部分中的预定的源区域,如例如图12B中所示。优选地,预定的目标区域非交叠,或者预定的源区域交叠。当预定的源区域是低于图3A的间隙填充起始频率309的第一组第一频谱部分的子集时,并且优选地,覆盖较低的频谱区域的预定的目标区域与具有间隙填充起始频率的其较低的频率边界重合,使得任何目标范围位于间隙填充起始频率之上,并且源范围位于间隙填充起始频率之下。

[0396] 如所讨论的,通过将目标区域与源区域相比较在目标区域与源区域之间不具有任意滞后但是同一目标区域具有特定滞后的情况下,获得精细粒度。在图11D的互相关性计算器1140中应用这些滞后,并且匹配对选择最终由铺片选择器1144执行。

[0397] 此外,优选的是执行块1142处示出的源范围和/或目标范围白化。然后,该块1142向比特流提供白化旗标,其用于控制图11B的解码器侧开关1123。此外,如果互相关性计算器1140提供负的结果,则该负的结果也被用信号通知给解码器。从而,在优选实施方式中,铺片选择器输出目标范围的源范围ID、滞后、符号,并且块1142额外地提供白化旗标。

[0398] 此外,参数计算器1134被配置成用于通过减少潜在的源范围的数目来执行源铺片修剪1146在于:基于相似度阈值从一组潜在的源铺片中除去源补片。从而,当两个源铺片之间的相似度大于或等于相似度阈值时,则这两个源铺片之一被从该组潜在的源中去除,并且该去除的源铺片不再用于进一步的处理,并且具体地,可以不由铺片选择器1144选择或者不用于块1140中执行的不同的源范围与目标范围之间的互相关性计算。

[0399] 已经参照不同的图描述了不同的实现。图1A至图5C涉及全速率或全带宽编码器/解码器方案。图6A至图7E涉及使用TNS或TTS处理的编码器/解码器方案。图8A至图8E涉及具有特定的双通道处理的编码器/解码器方案。图9A至图10D涉及特定的能量信息计算和应用,以及图11A至图12C涉及铺片选择的具体方式。

[0400] 所有这些不同的方面可以是相互无关的本发明的用途,但是另外地,也可以基本上如图2A和图2B中所示,一起被应用。然而,特定的双通道处理也可以被应用于如图13中所

示的编码器/解码器方案,并且对于TNS/TTS处理、重构频带中的包络能量信息计算和应用或自适应源范围标识和解码器侧的相应的应用同样成立。另一方面,在使用或不使用TNS/TTS处理,使用或不使用双通道处理,使用或不使用自适应源范围标识或使用频谱包络表示的其他类型的能量计算的情况下,全速率方面适用。从而,清楚的是,这些各个方面之一的特征也可以适用于其他方面中。

[0401] 虽然在用于编码或解码的设备的背景下描述了一些方面,但是清楚的是,这些方面还给出相应的方法的描述,其中,一个块或装置对应于方法步骤或者方法步骤的特征。类似地,在方法步骤的背景下描述的方面还给出相应的块或项目的描述或者相应的设备的特征。一些或全部方法步骤可以由(或者使用)硬件设备例如微处理器、可编码计算机或电子电路来执行。在一些实施方式中,最重要的方法步骤中的某些一个或更多个可以由这样的设备来执行。

[0402] 取决于某些实现要求,本发明的实施方式可以以硬件或软件被实现。实现可以使用例如数字存储介质的非暂态存储介质来实现,例如软盘、硬盘驱动(HDD)、DVD、蓝光、CD、ROM、PROM和EPROM、EEPROM或者闪存存储器,其上存储有电可读的控制信号,其与可编程计算机系统合作(或能够与其合作),使得执行相应的方法。因此,数字存储介质可以是计算机可读的。

[0403] 根据本发明的一些实施方式包括数据载体,数据载体具有电子可读控制信号,电子可读控制信号能够与可编程计算机系统协作以使得执行本文中描述的方法之一。

[0404] 通常,本发明的实施方式可以被实现为具有程序代码的计算机程序产品,当计算机程序产品在计算机上运行时,程序代码操作地用于执行方法之一。程序代码可以被存储在例如机器可读载体上。

[0405] 其他实施方式包括被存储在机器可读载体上的用于执行本文中描述的方法之一的计算机程序。

[0406] 换言之,本发明的方法的实施方式因此是具有程序代码的计算机程序,当计算机程序在计算机上运行时,程序代码用于执行本文中描述的方法之一。

[0407] 因此,本发明的方法的又一实施方式是数据载体(或数字存储介质或者计算机可读介质),该数据载体包括其上存储的用于执行本文中描述的方法之一的计算机程序。数据载体、数字存储介质或记录介质通常为有形的和/或非暂态。

[0408] 因此,本发明的方法的另一实施方式是表示用于执行本文中描述的方法之一的计算机程序的数据流或信号序列。数据流或信号序列可以被配置成例如经由数据通信连接例如经由因特网被传送。

[0409] 又一实施方式包括处理装置例如计算机或可编程逻辑器件,处理装置被配置成或适于执行本文中描述的方法之一。

[0410] 又一实施方式包括其上安装有用于执行本文中描述的方法之一的计算机程序的计算机。

[0411] 根据本发明的又一实施方式包括被配置成将用于执行本文中描述的方法之一的计算机程序(例如,电子地或光学地)传送至接收器的设备或系统。接收器可以是例如计算机、移动装置、存储装置等。设备或系统可以包括例如用于将计算机程序传送至接收器的文件服务器。

[0412] 在一些实施方式中,可编程逻辑器件(例如,现场可编程门阵列)可以用于执行本文中描述的方法的一些或全部功能。在一些实施方式中,现场可编程门阵列可以与微处理器协作以执行本文中描述的方法之一。通常,优选地,这些方法由任意硬件设备来执行。

[0413] 上述实施方式对于本发明的原理仅为说明性。应当理解,对本领域中的其他技术人员而言,本文中描述的布置和细节的修改和变化会很明显。因此,意在仅受所附专利权利要求的范围的限制,而不受借助于本文中的实施方式的描述和说明给出的具体细节的限制。

[0414] 引用文献列表

[0415] [1]Dietz、L.Liljeryd、K.**Kjörling**和O.Kunz,“Spectral Band Replication,a novel approach in audio coding”,第112次AES(音频工程协会)会议,慕尼黑,2002年5月

[0416] [2]Ferreira、D.Sinha,“Accurate Spectral Replacement”,音频工程协会会议,巴塞罗那,西班牙,2005年

[0417] [3]D.Sinha、A.Ferreiral和E.Harinarayanan,“A Novel Integrated Audio Bandwidth Extension Toolkit (ABET)”,音频工程协会会议,巴黎,法国,2006年

[0418] [4]R.Annadana、E.Harinarayanan、A.Ferreira和D.Sinha,“New Results in Low Bit Rate Speech Coding and Bandwidth Extension”,音频工程协会会议,旧金山,美国,2006年

[0419] [5]T.**Żernicki**、M.Bartkowiak,“Audio bandwidth extension by frequency scaling of sinusoidal partials”,音频工程协会会议,旧金山,美国,2008年

[0420] [6]J.Herre、D.Schulz,Extending the MPEG-4AAC Codec by Perceptual Noise Substitution,第104次AES(音频工程协会)会议,阿姆斯特丹,1998年,预印本4720

[0421] [7]M.Neuendorf、M.Multrus、N.Rettelbach等人,MPEG Unified Speech and Audio Coding-The ISO/MPEGStandard for High-Efficiency Audio Coding of all Content Types,第132次AES(音频工程协会)会议,布达佩斯,匈牙利,2012年4月

[0422] [8]McAulay、Robert J.、Quatieri、Thomas F,“Speech Analysis/Synthesis Based on a Sinusoidal Representation”,关于声学、语音和信号处理的IEEE会报,第34(4)卷,1986年8月

[0423] [9]Smith,J.O.、Serra,X.,“PARSHL:An analysis/synthesis program for non-harmonic sounds based on a sinusoidal representation”,国际计算机音乐会议的会议记录,1987年

[0424] [10]Purnhagen、H.;Meine、Nikolaus,“HILN-the MPEG-4parametric audio coding tools”,电路与系统,2000年ISCAS(电路与系统国际研讨会)会议记录,日内瓦,2000年,2000IEEE国际研讨会,第3卷,第201、204页,第3卷,2000

[0425] [11]国际标准ISO/IEC 13818-3,“Generic Coding of Moving Pictures and Associated Audio:Audio”,日内瓦,1998年

[0426] [12]M.Bosi、K.Brandenburg、S.Quackenbush、L.Fielder、K.Akagiri、H.Fuchs、M.Dietz、J.Herre、G.Davidson、Oikawa,“MPEG-2Advanced Audio Coding”,第101次AES(音频工程协会)会议,洛杉矶,1996年

[0427] [13]J.Herre,“Temporal Noise Shaping,Quantization and Coding methods

in Perceptual Audio Coding: A Tutorial introduction”,关于高质量音频编码的第17次AES((音频工程协会)国际会议,1999年8月

[0428] [14] J. Herre, “Temporal Noise Shaping, Quantization and Coding methods in Perceptual Audio Coding: A Tutorial introduction”,关于高质量音频编码的第17次AES((音频工程协会)国际会议,1999年8月

[0429] [15] 国际标准ISO/IEC 23001-3:2010, Unified speech and audio coding Audio, 日内瓦, 2010年

[0430] [16] 国际标准ISO/IEC 14496-3:2005, Information technology-Coding of audio-visual objects-Part 3: Audio, 日内瓦, 2005年

[0431] [17] P. Ekstrand, “Bandwidth Extension of Audio Signals by Spectral Band Replication”,关于MPCA的第1次IEEE比荷卢经济联盟研讨会的会议记录, 勒芬, 2002年11月

[0432] [18] F. Nagel, S. Disch, S. Wilde, A continuous modulated single sideband bandwidth extension,关于声学、语音和信号处理的ICASSP国际会议, 达拉斯, 德克萨斯州(美国), 2010年4月

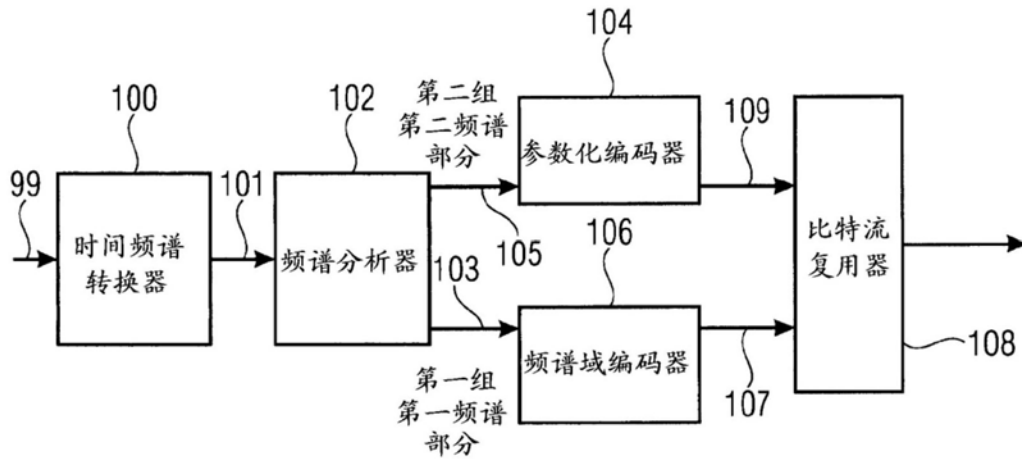


图1A

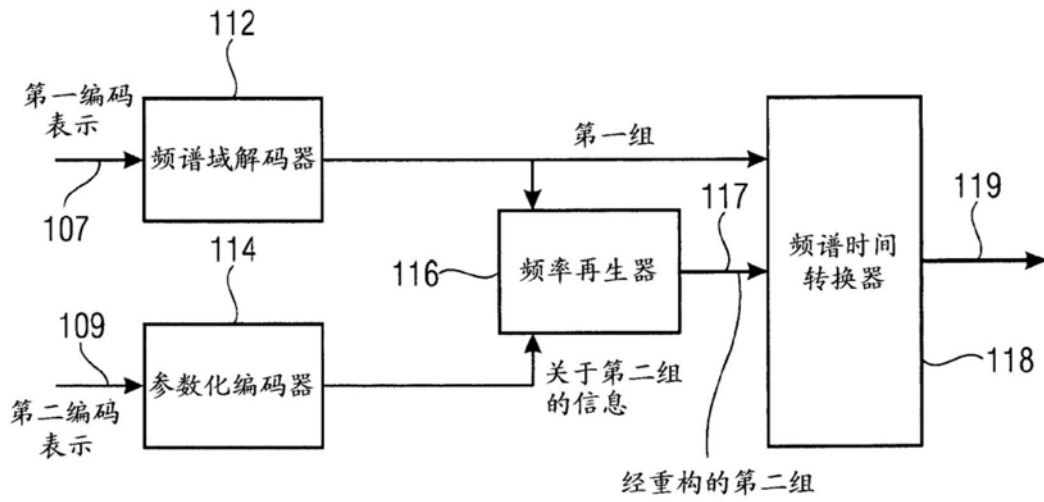


图1B

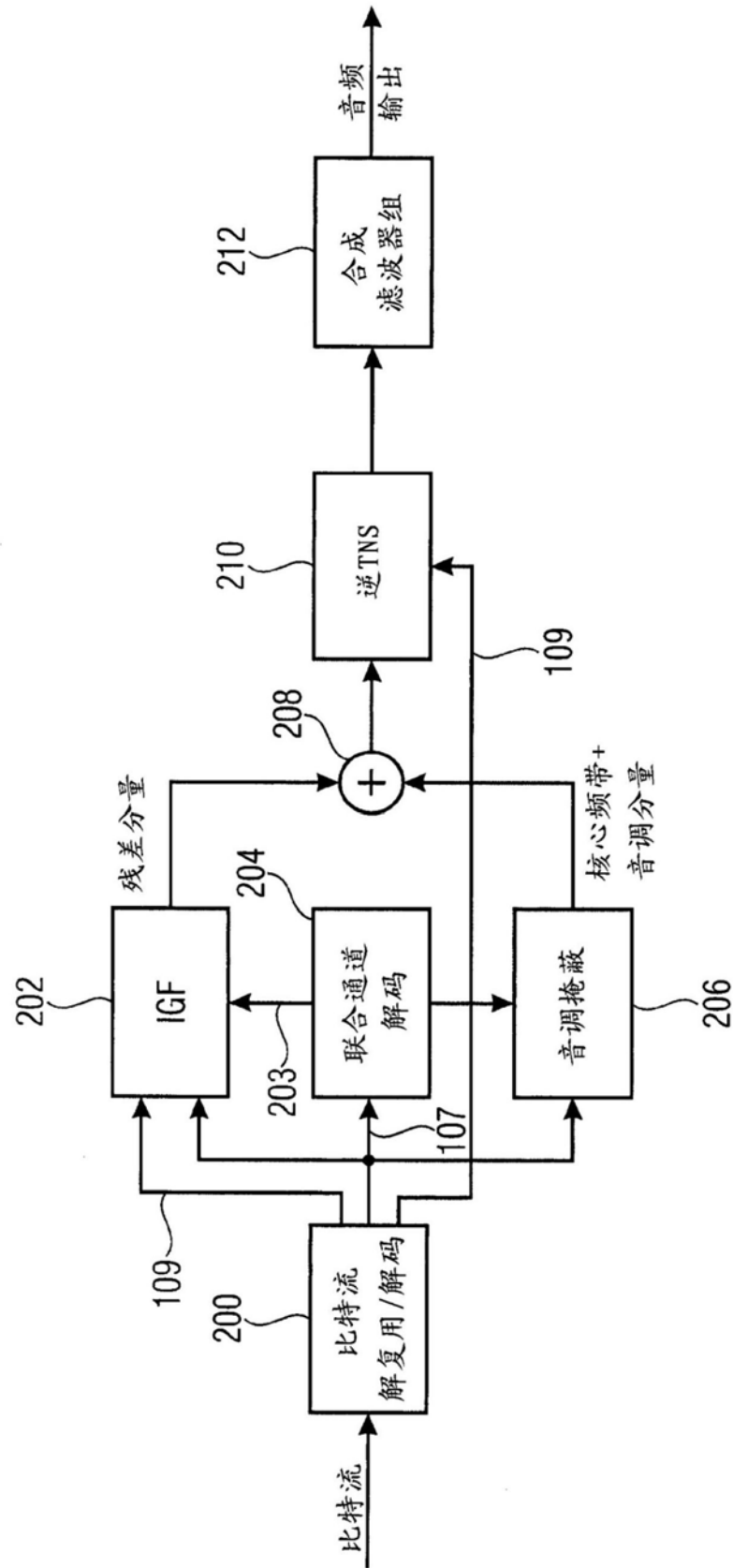


图2A

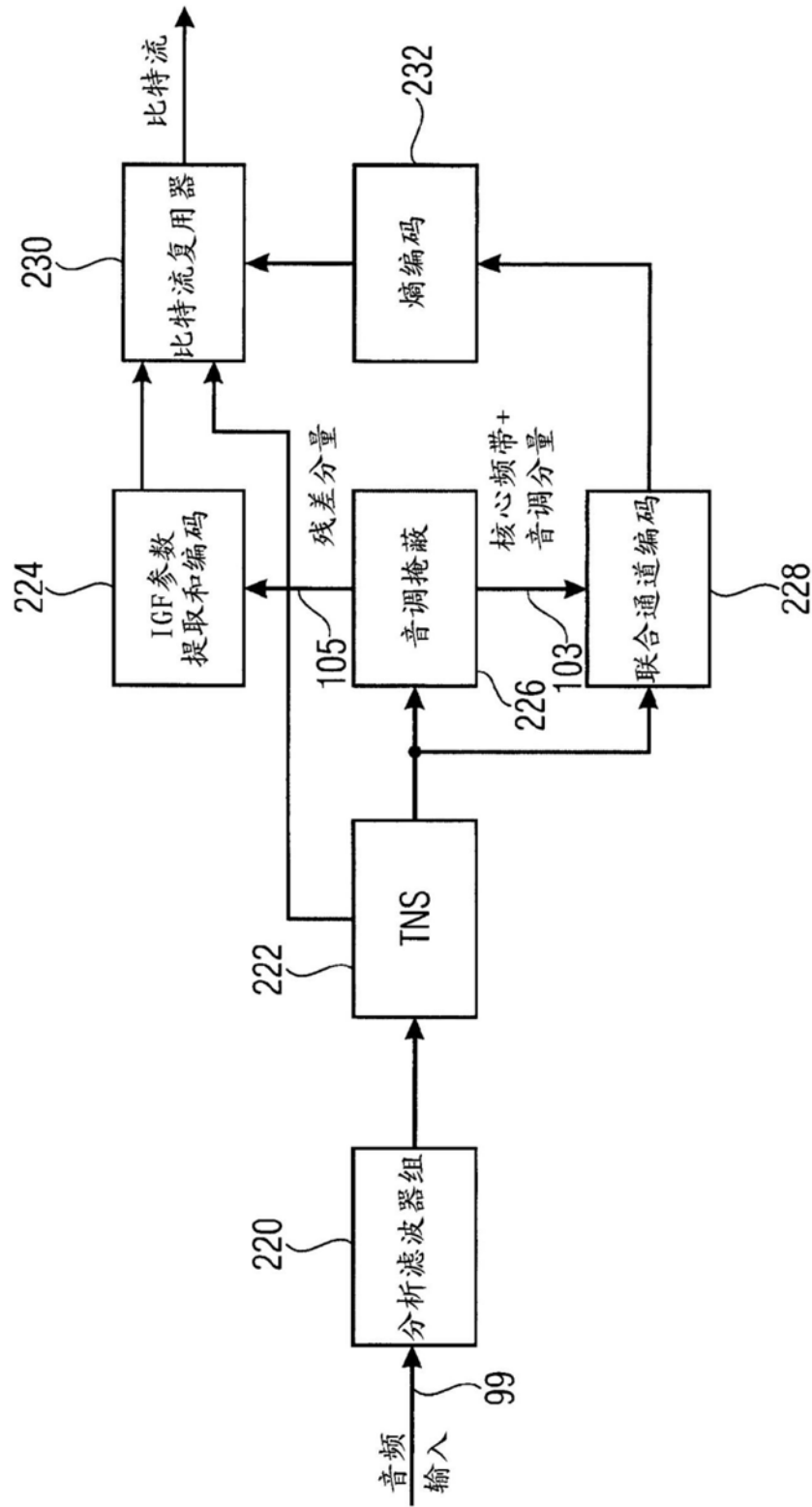


图2B

- 用于第一组的“包络”
 (线状编码)的第一分辨率(高分辨率)
- 用于第二组的“包络”
 (每SCB的比例因子)的第二分辨率(低分辨率)

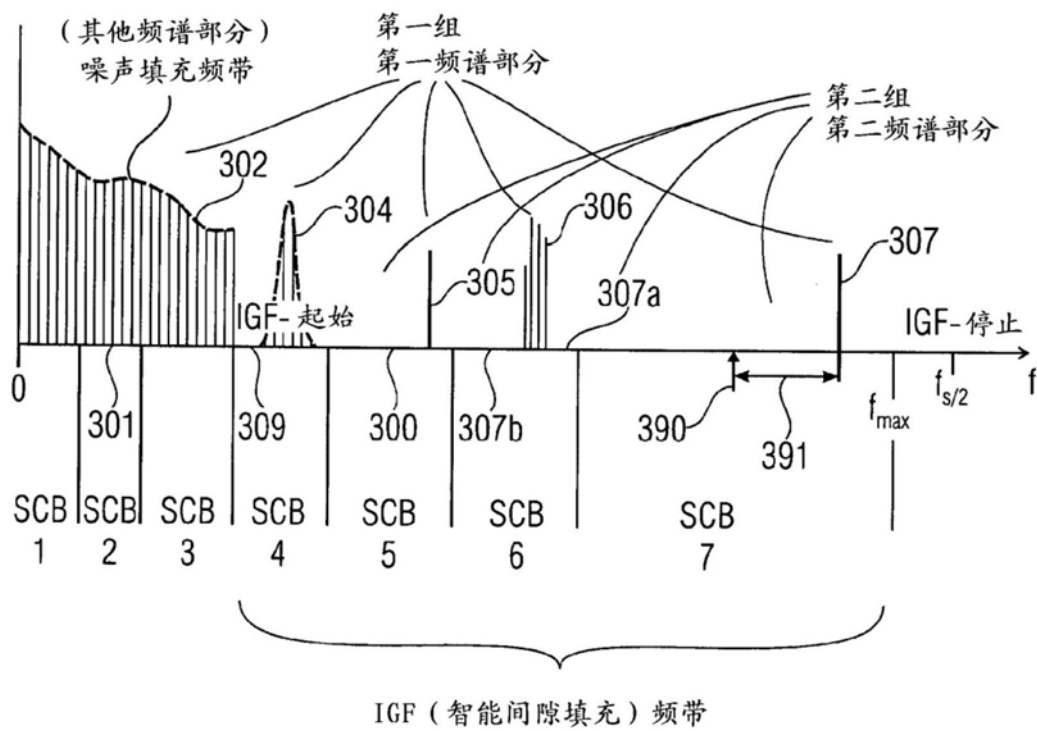


图3A

SCB1	SCB2	SCB3	SCB4	SCB5	SCB6	SCB7
SF1	SF2	SF3	SF4	SF5	SF6	SF7
			E ₁	E ₂	E ₃	E ₄
	NF ₂					

308

310

312

图3B

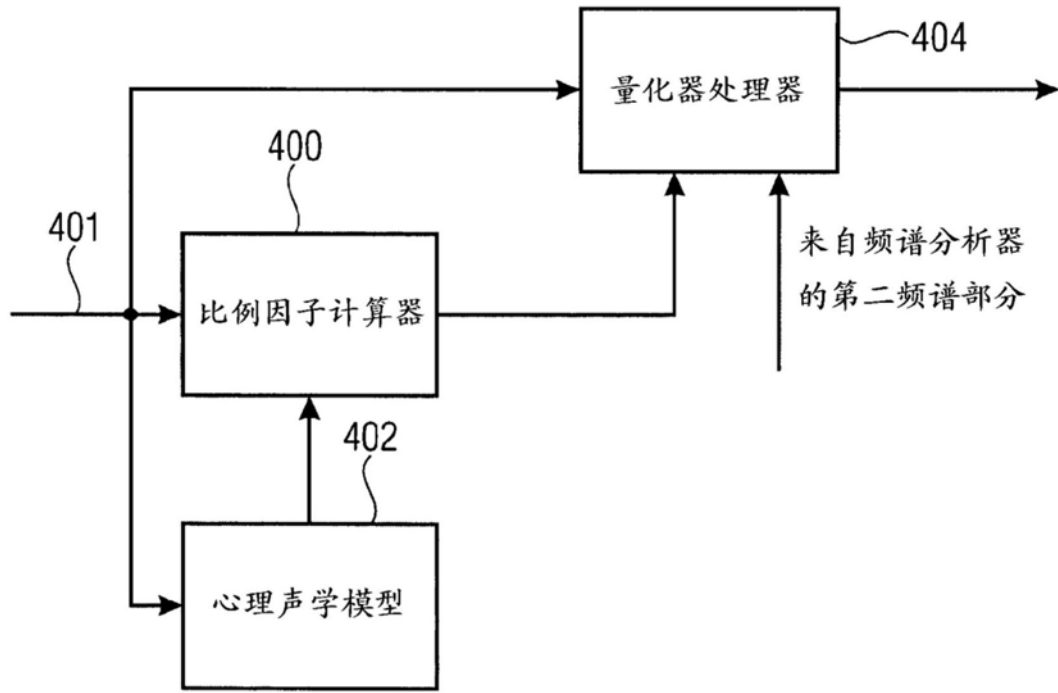


图4A

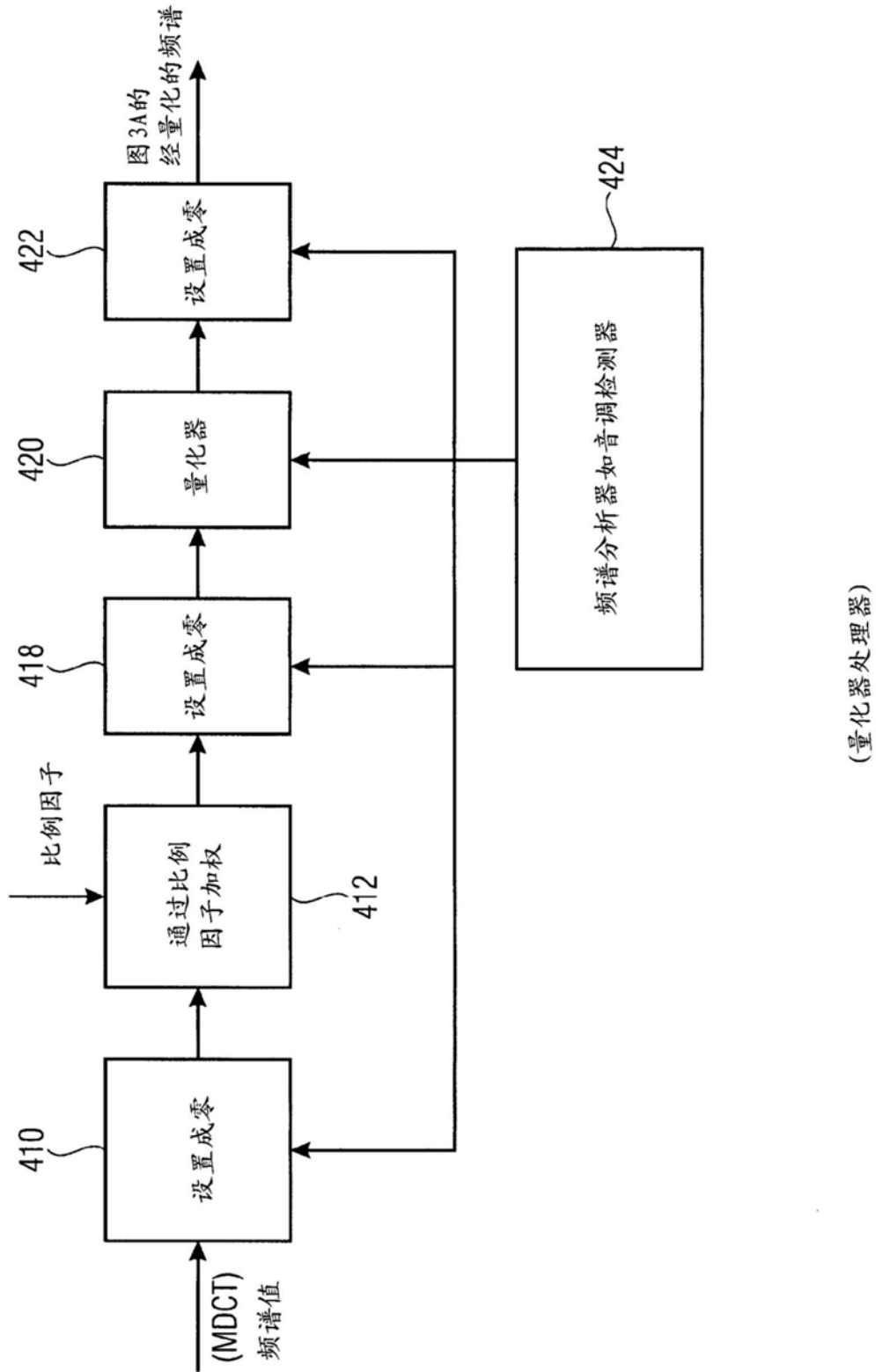


图4B

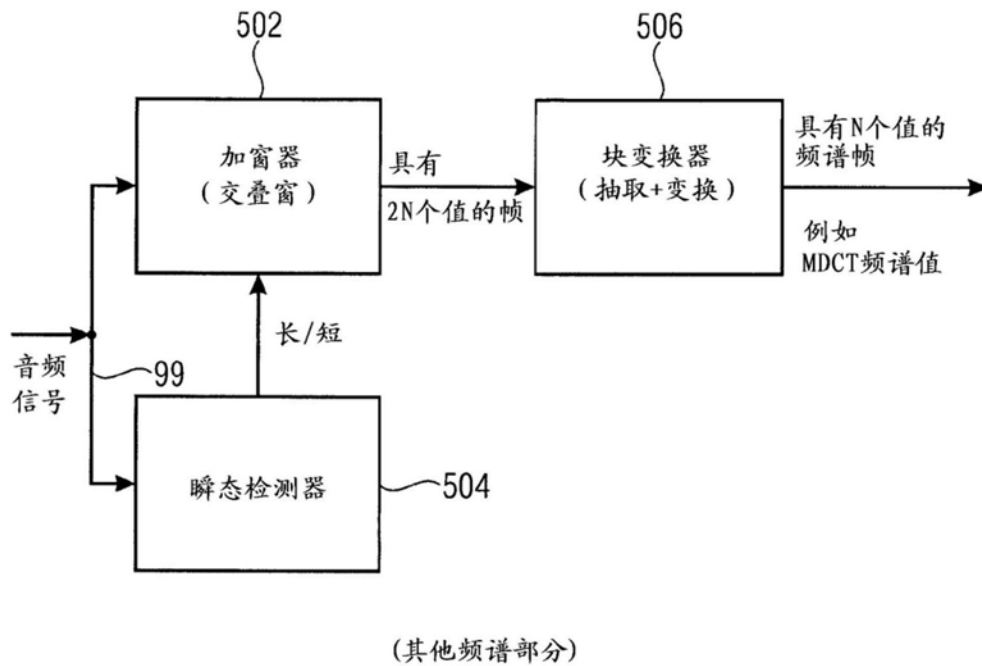


图5A

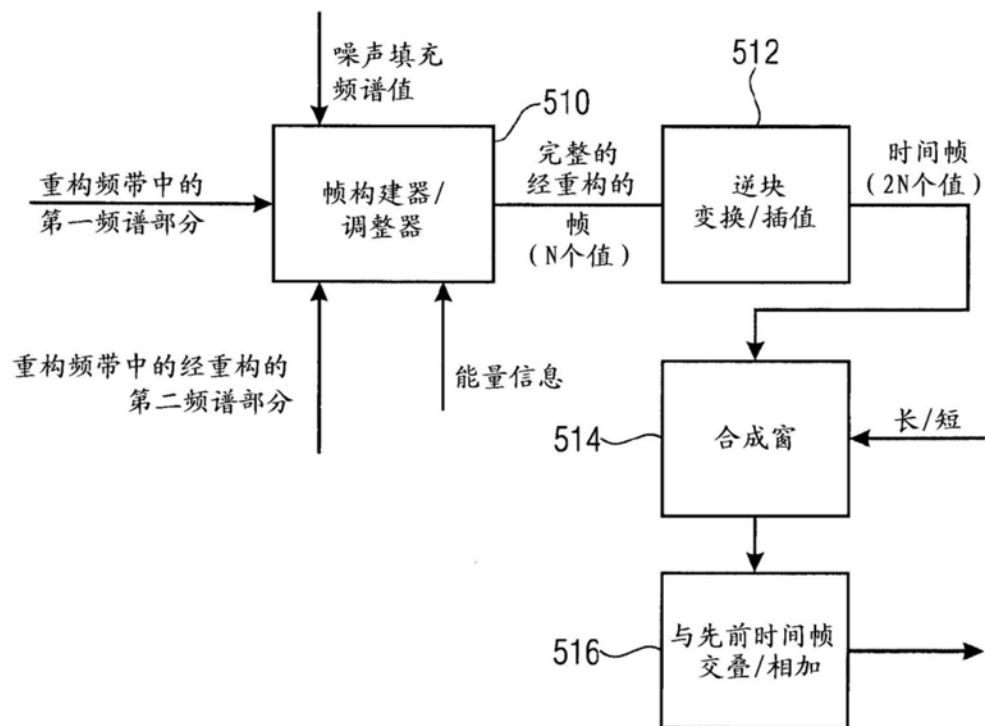


图5B

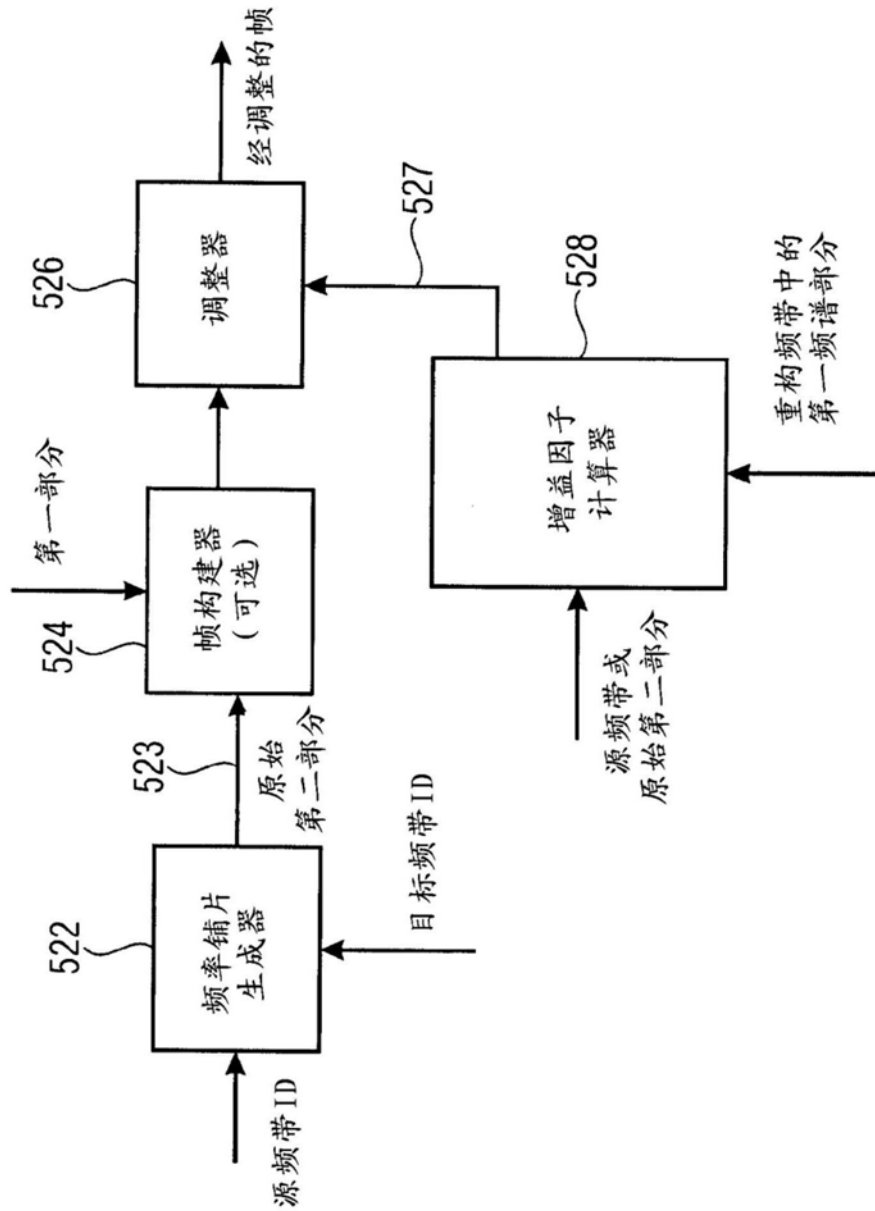


图5C

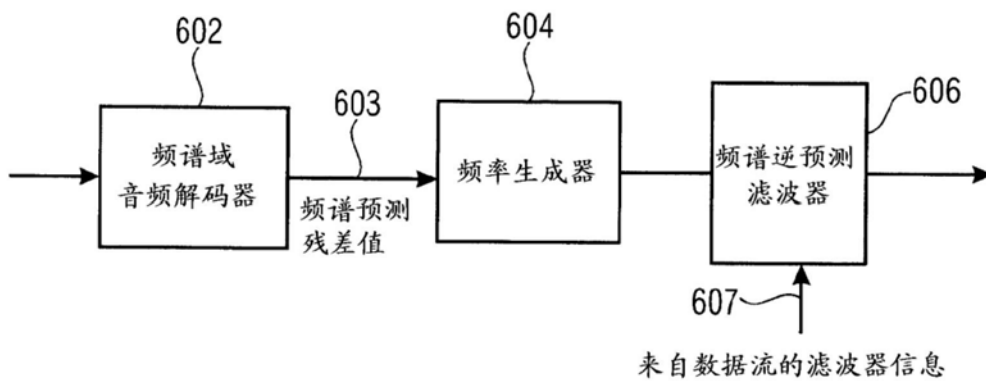


图6A

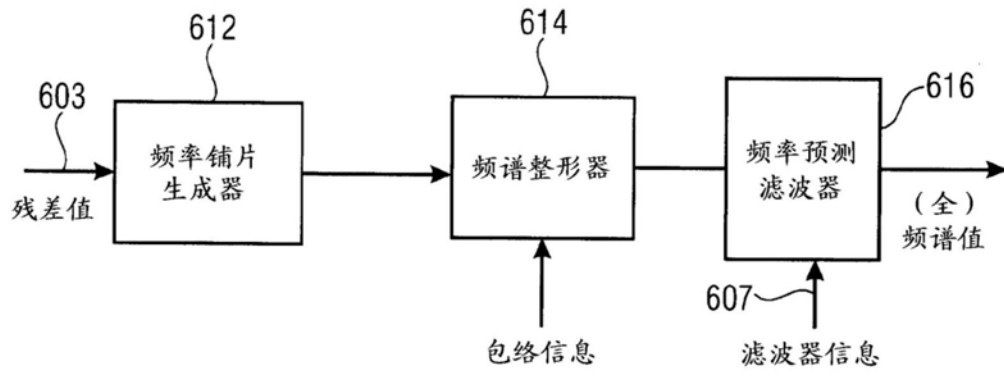


图6B

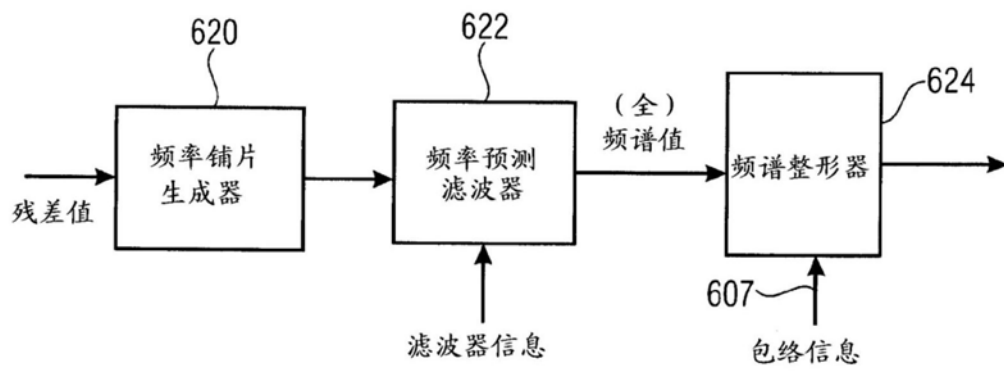


图6C

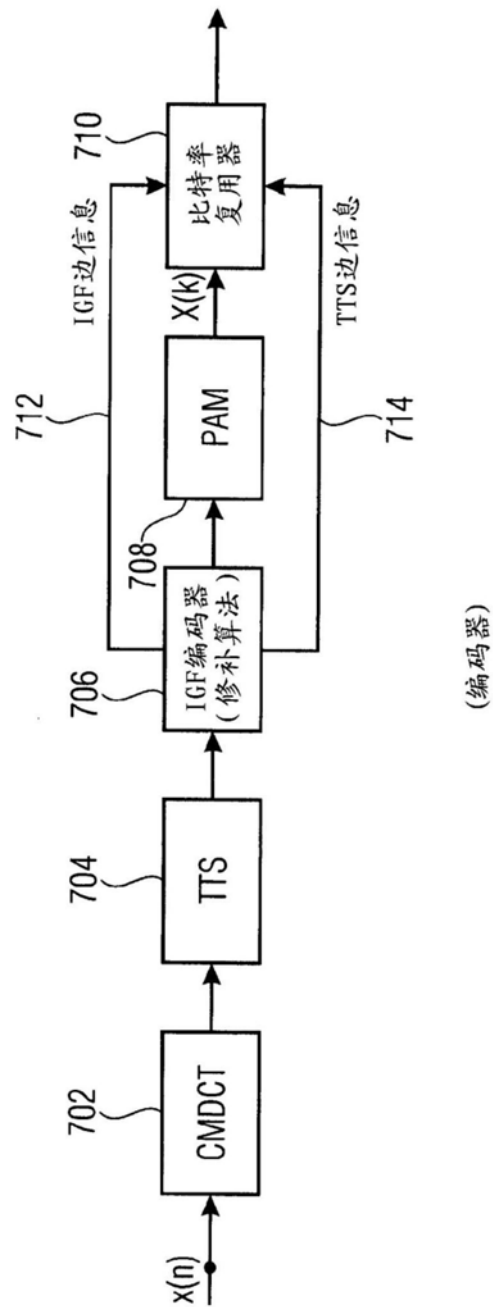


图7A

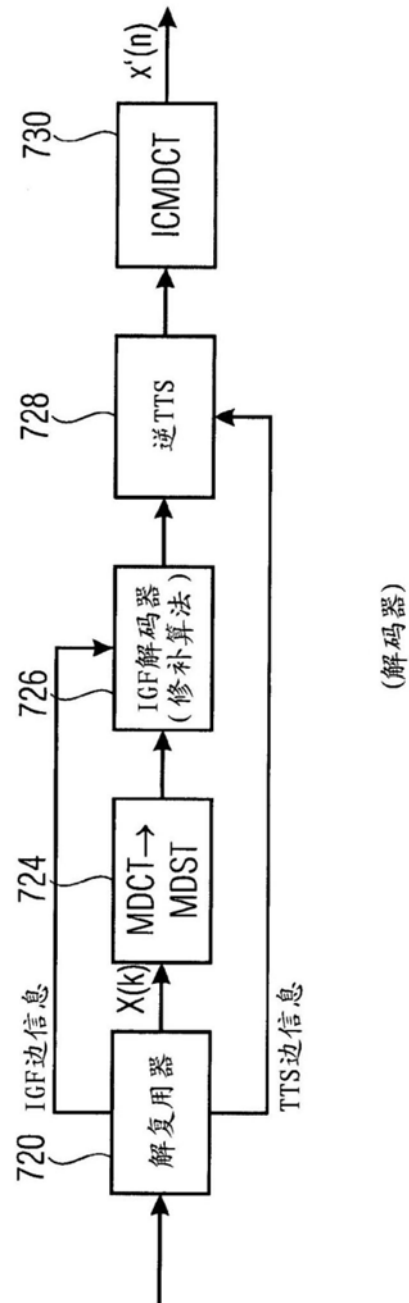


图7B

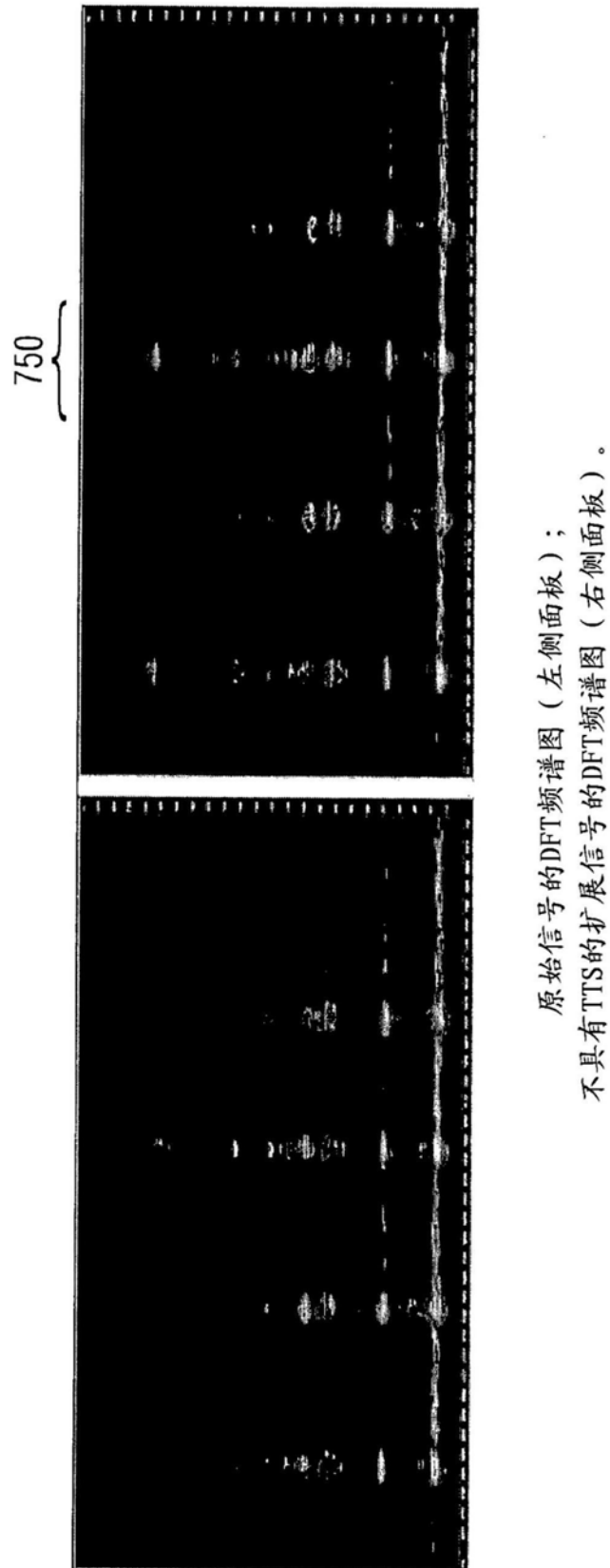
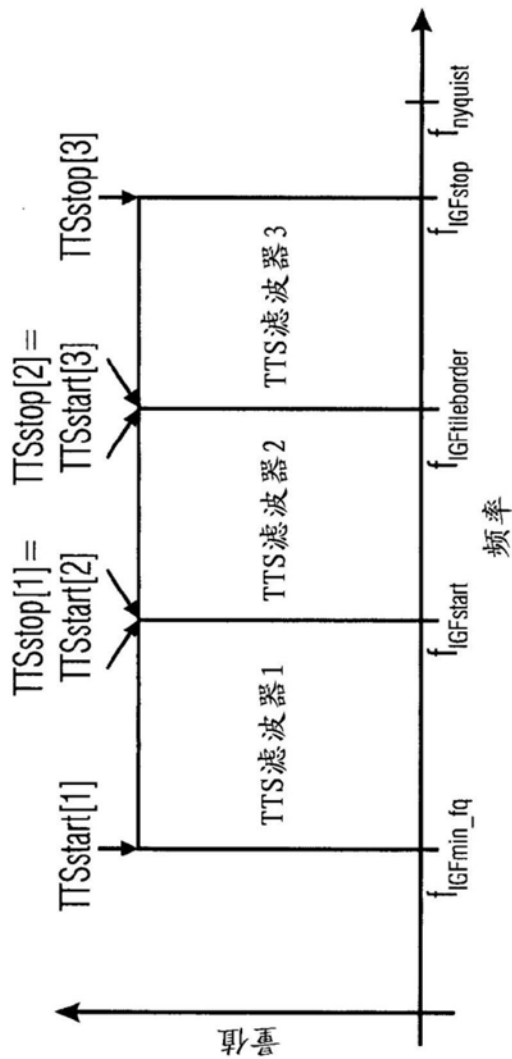


图7C



基本的TTS滤波器方案。每个框标记由一个特别的滤波器所覆盖的频谱区域。

图7D

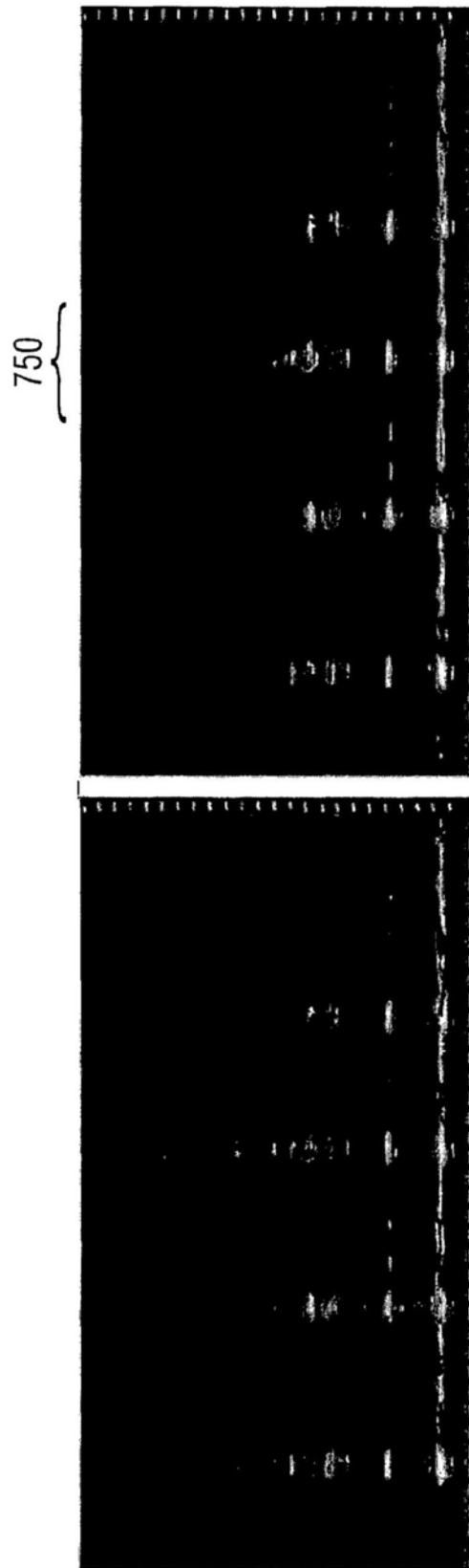


图7E

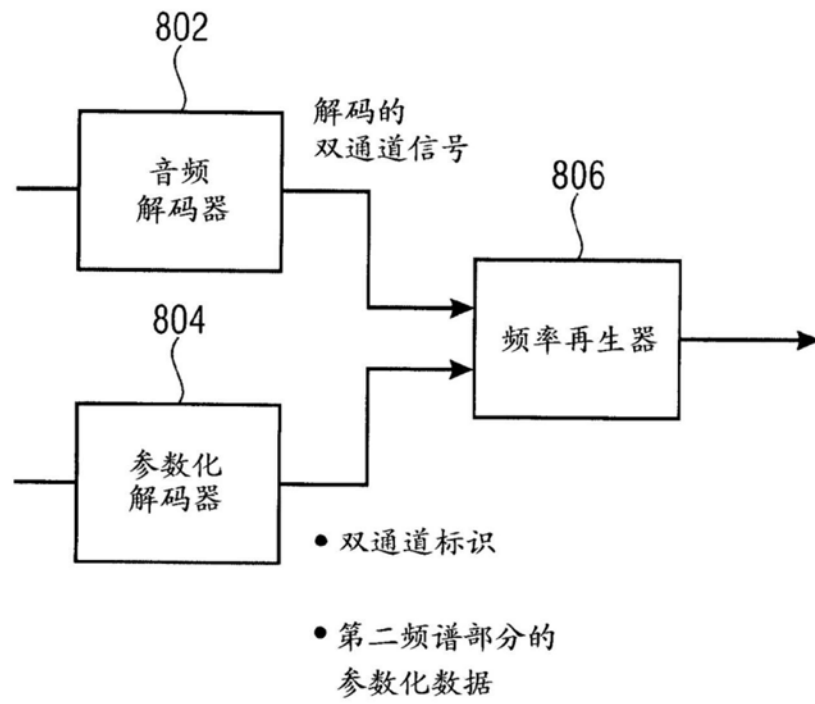


图8A

源范围	1	1	2	2	1: 第一双通道 表示 (例如L/R)
目标范围	1	2	1	2	

2: 第二双通道
表示 (例如M/S)

图8B

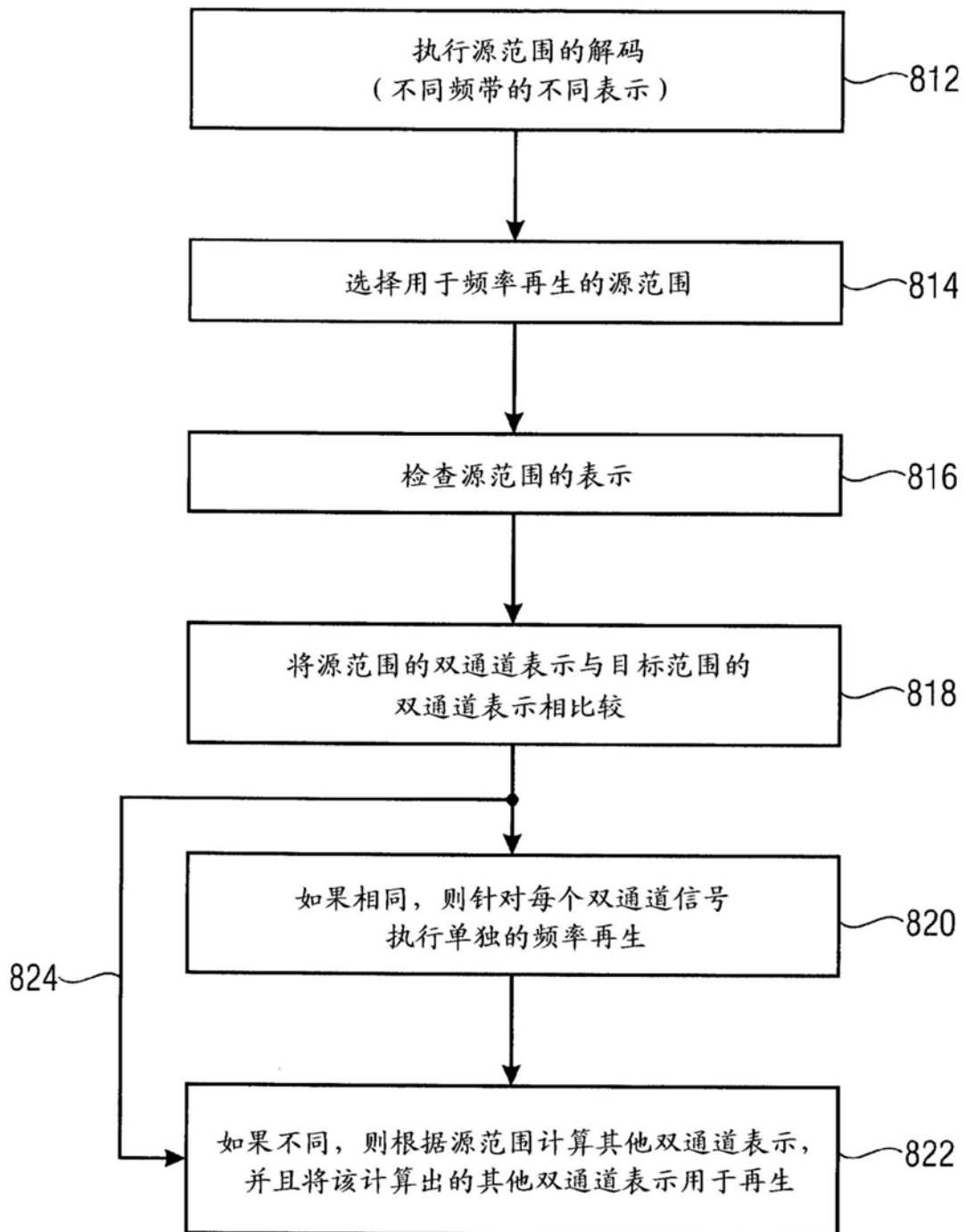


图8C

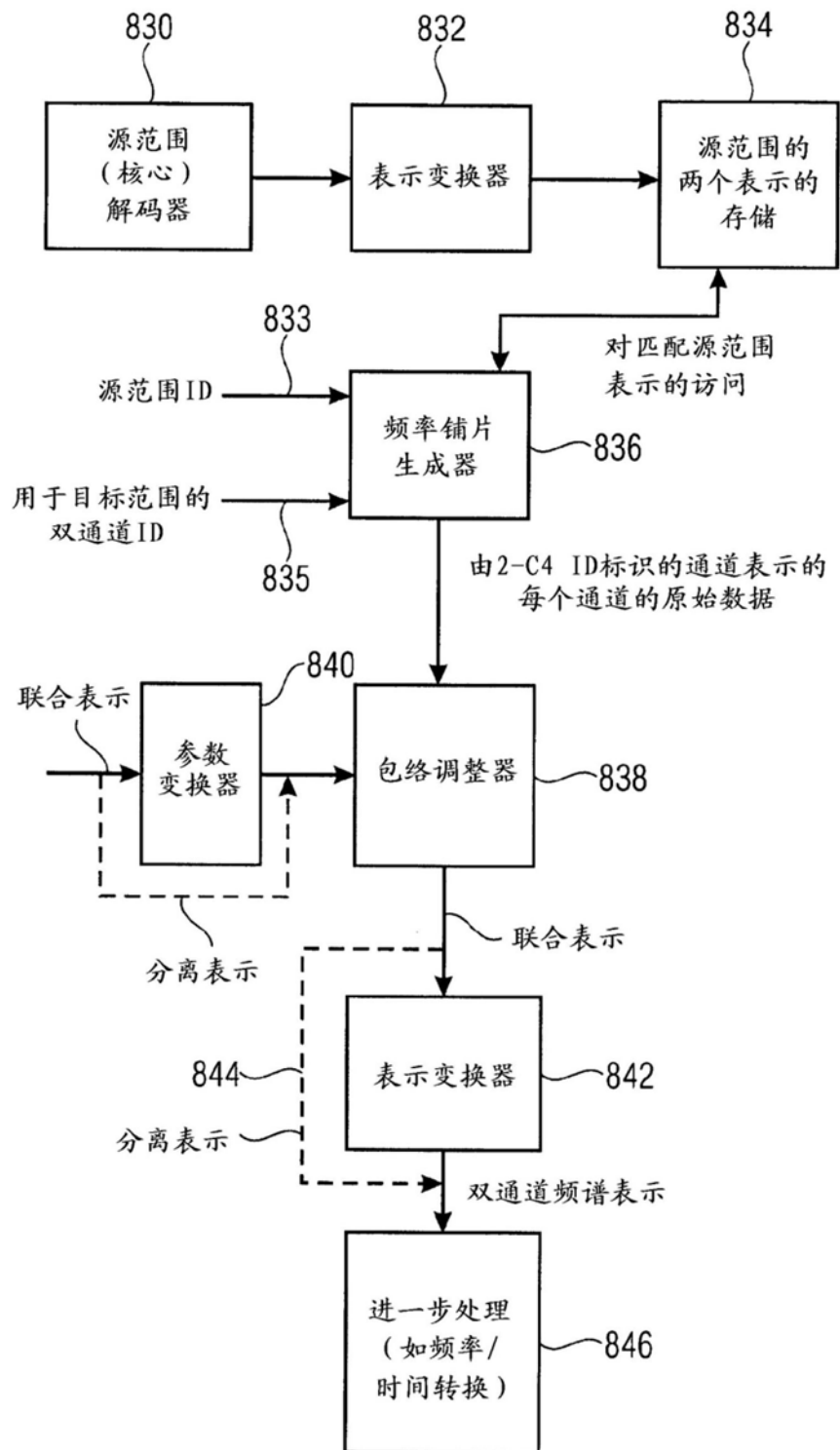


图8D

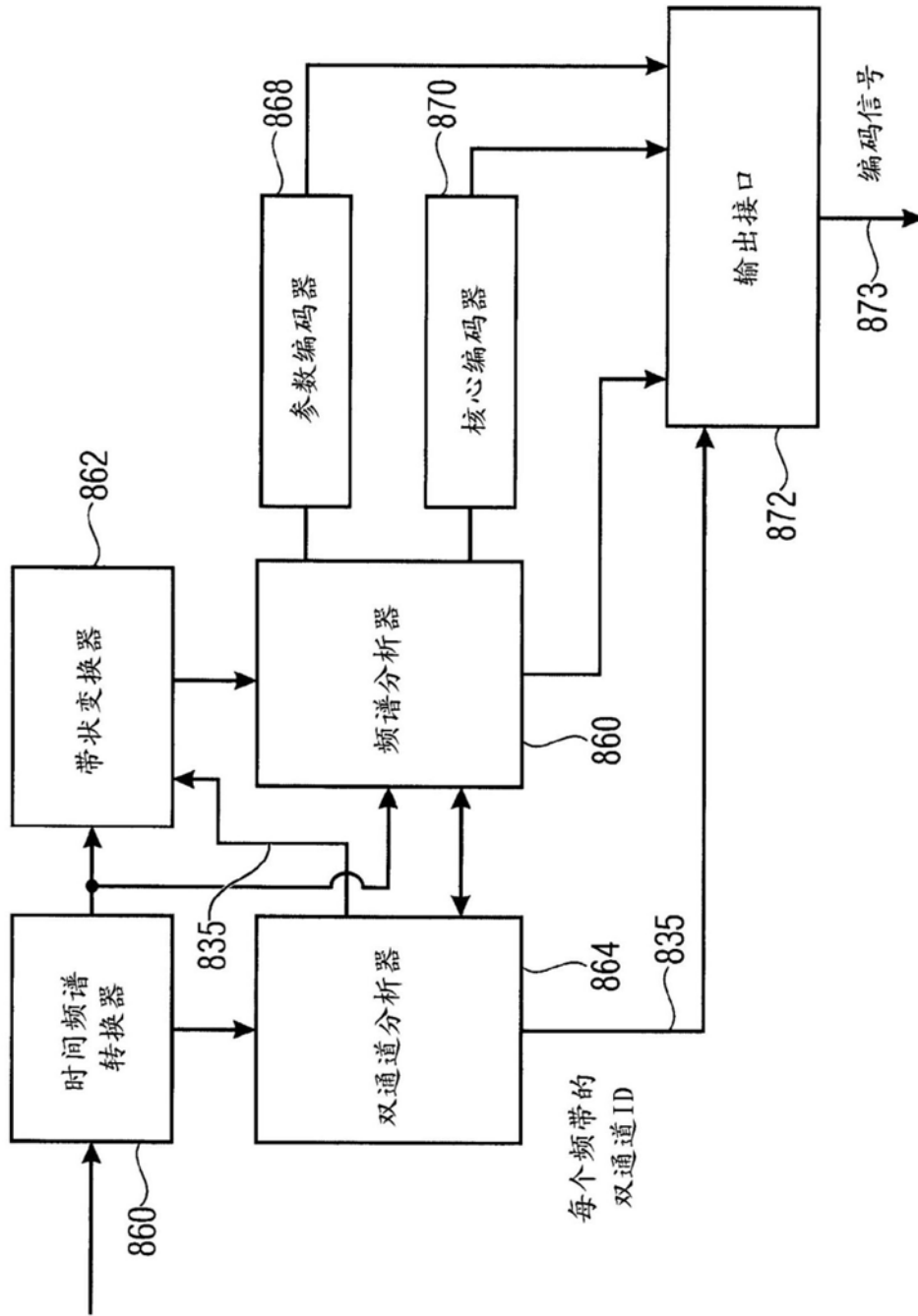


图8E

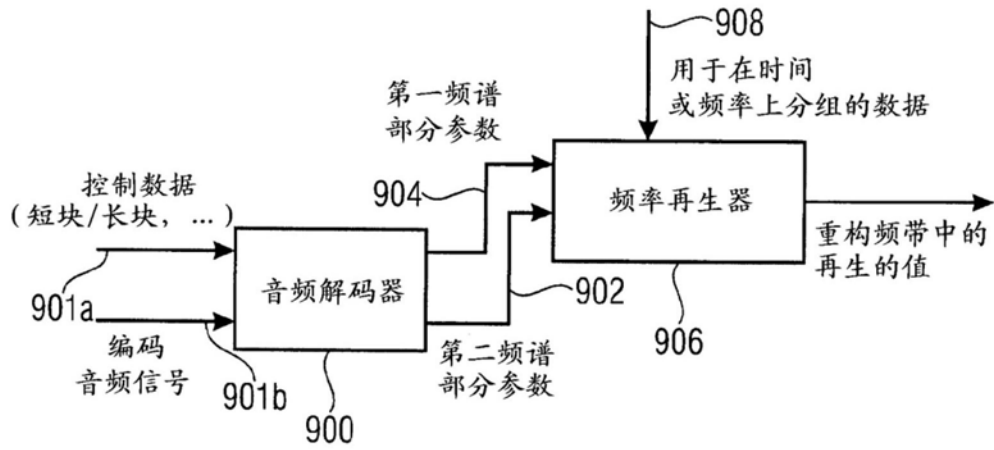


图9A

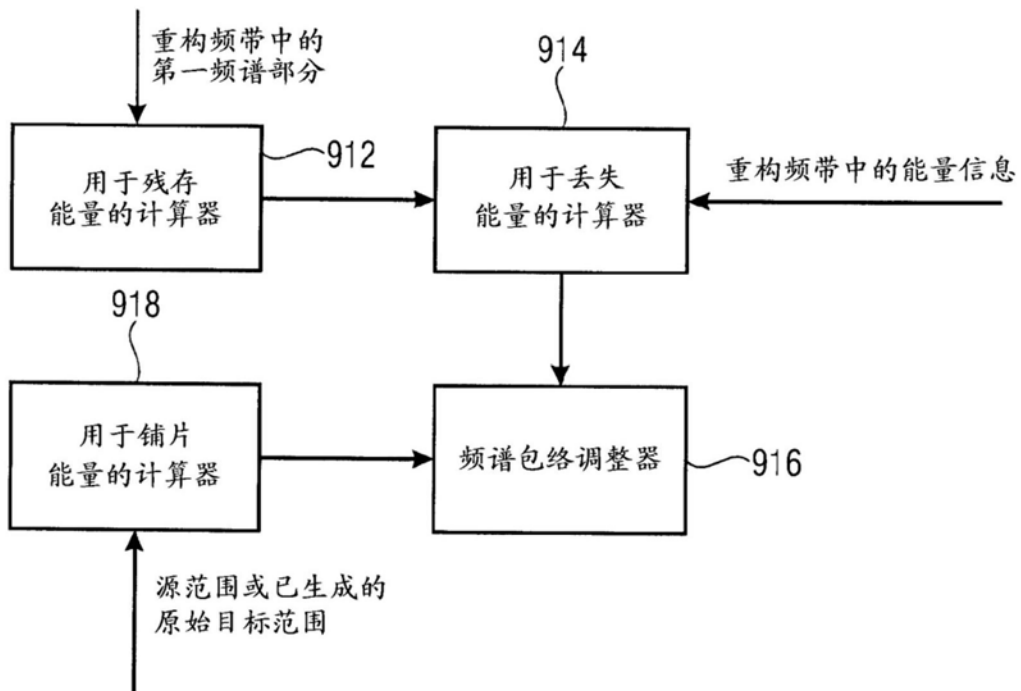
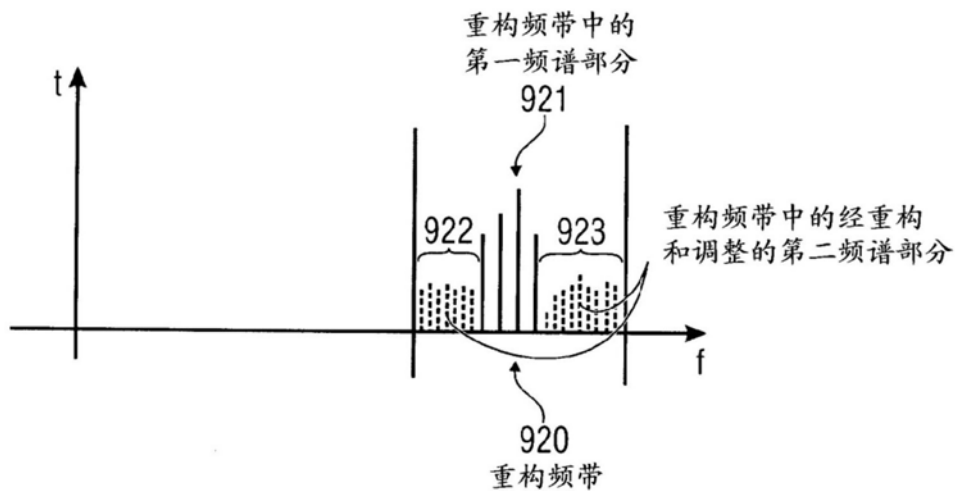


图9B



- 例如 ● 残存能量: 5单位
- 重构频带的能量值: 5单位 (覆盖重构频带中的
第一和第二频谱部分)
- 源范围数据或原始
目标范围数据的能量: 8单位
- 丢失的能量: 5单位
- 增益因子: $g := \sqrt{\frac{mE_k}{pE_k}} = 0.79$

→ 仅对第二频谱部分的频谱值进行调整

→ 第一频谱部分不受包络调整影响

图9C

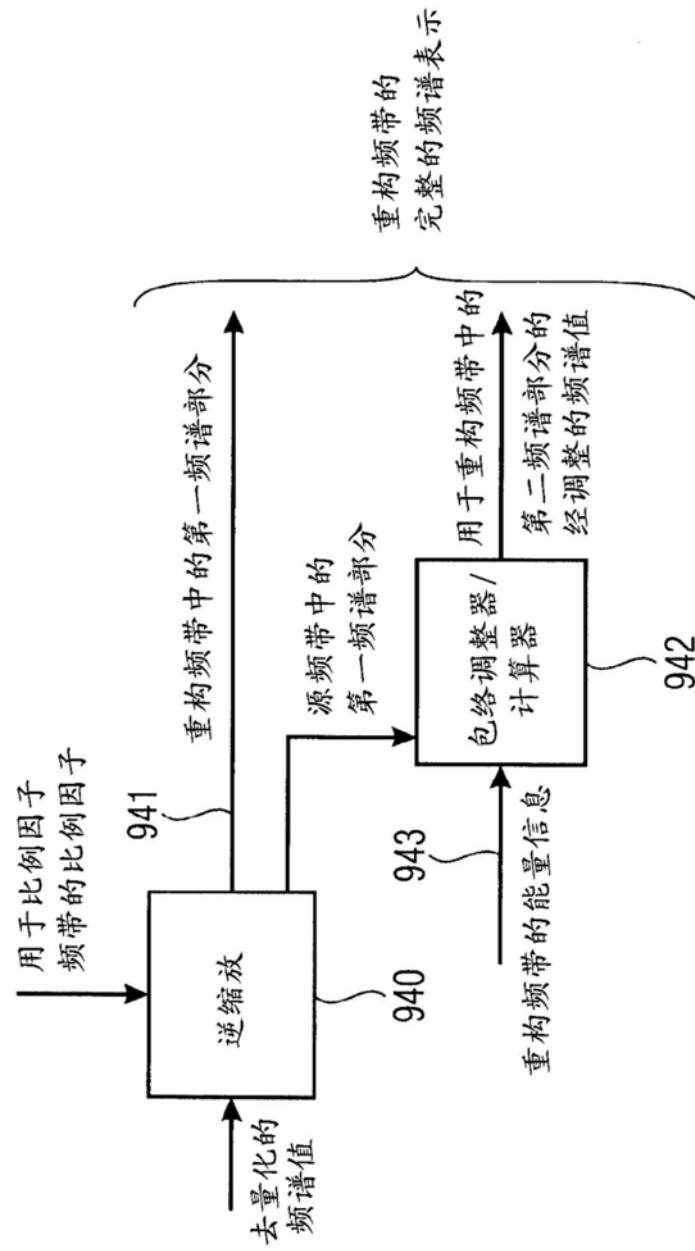


图9D

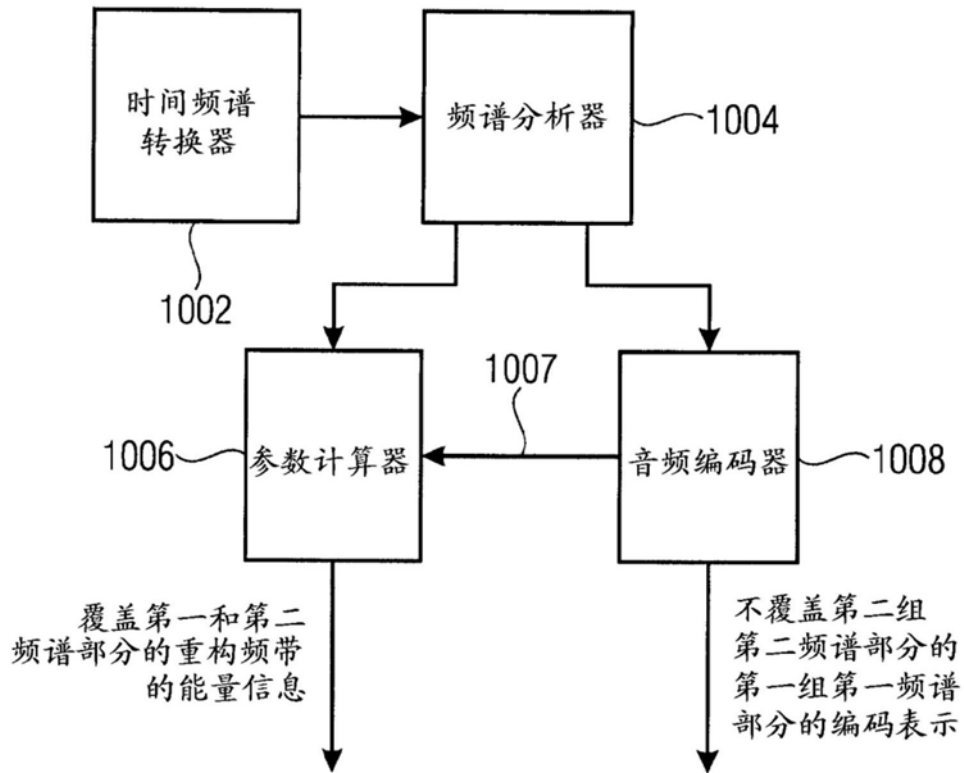


图10A

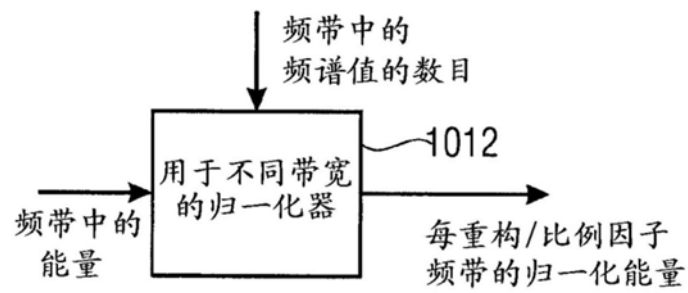


图10B

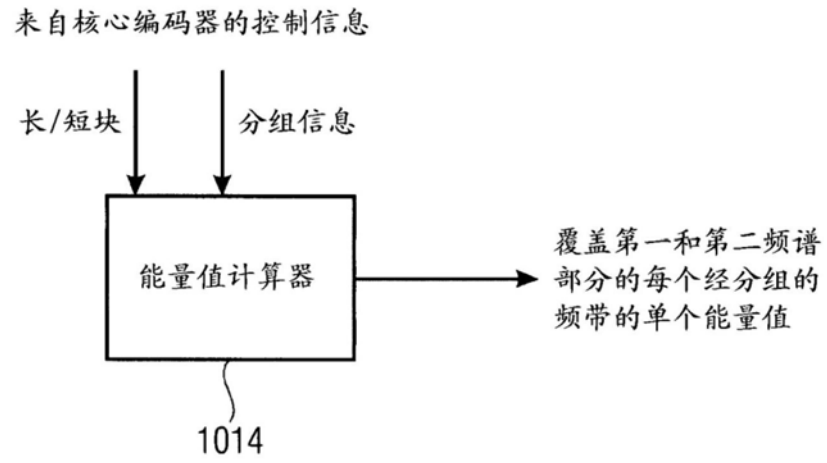


图10C

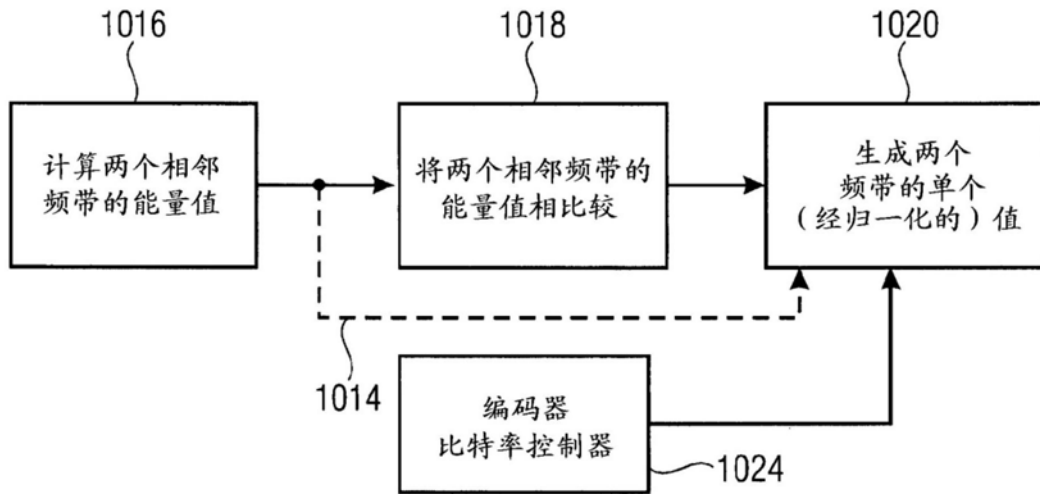


图10D

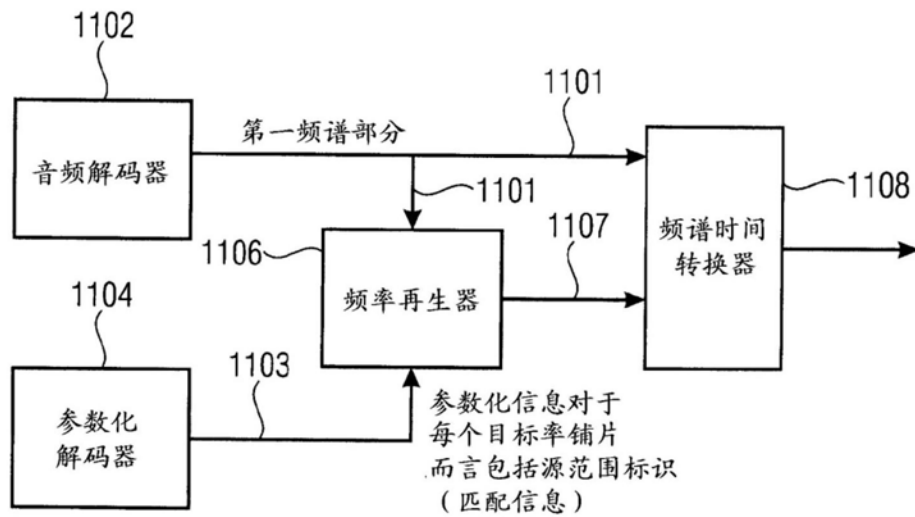


图11A

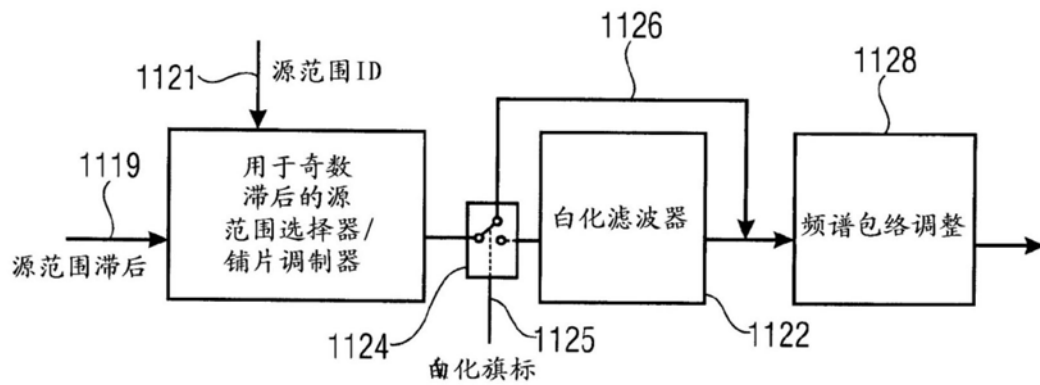


图11B

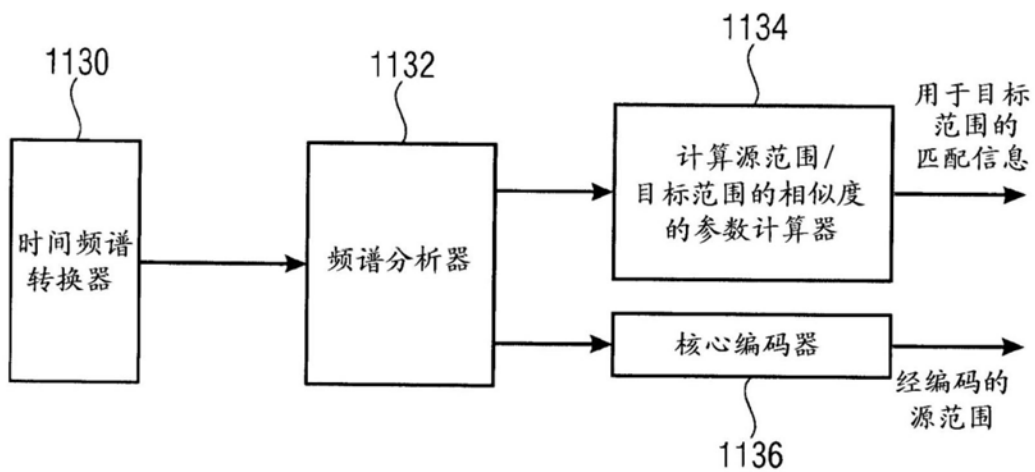


图11C

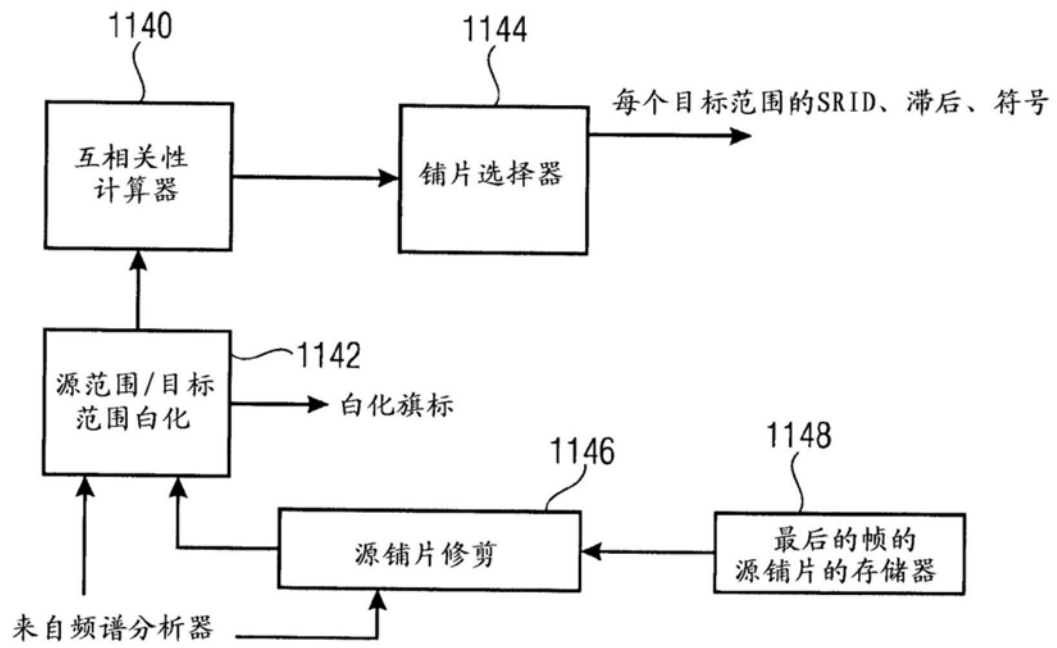


图11D

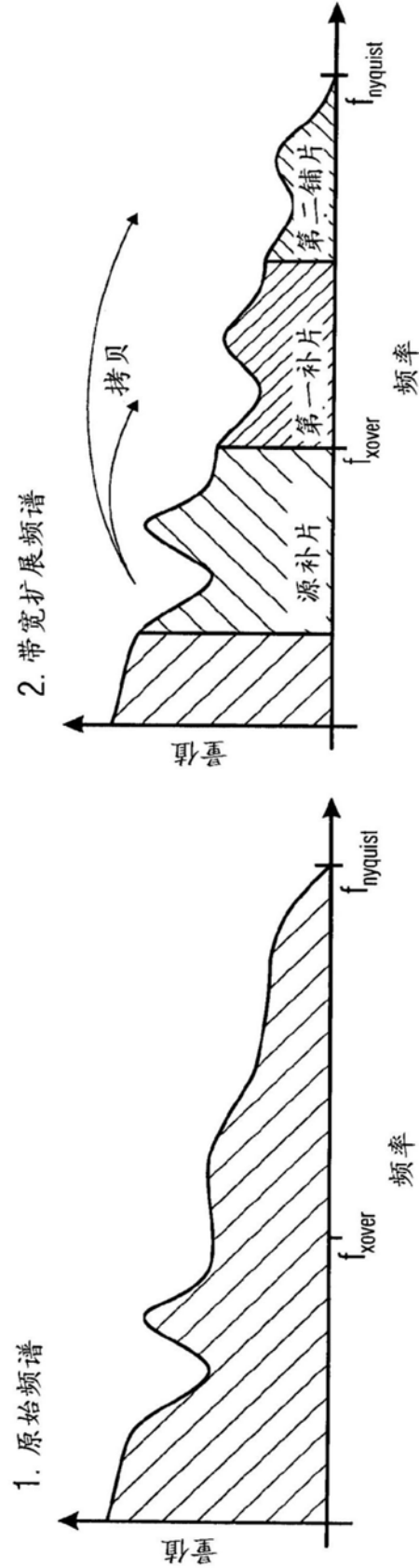


图12A

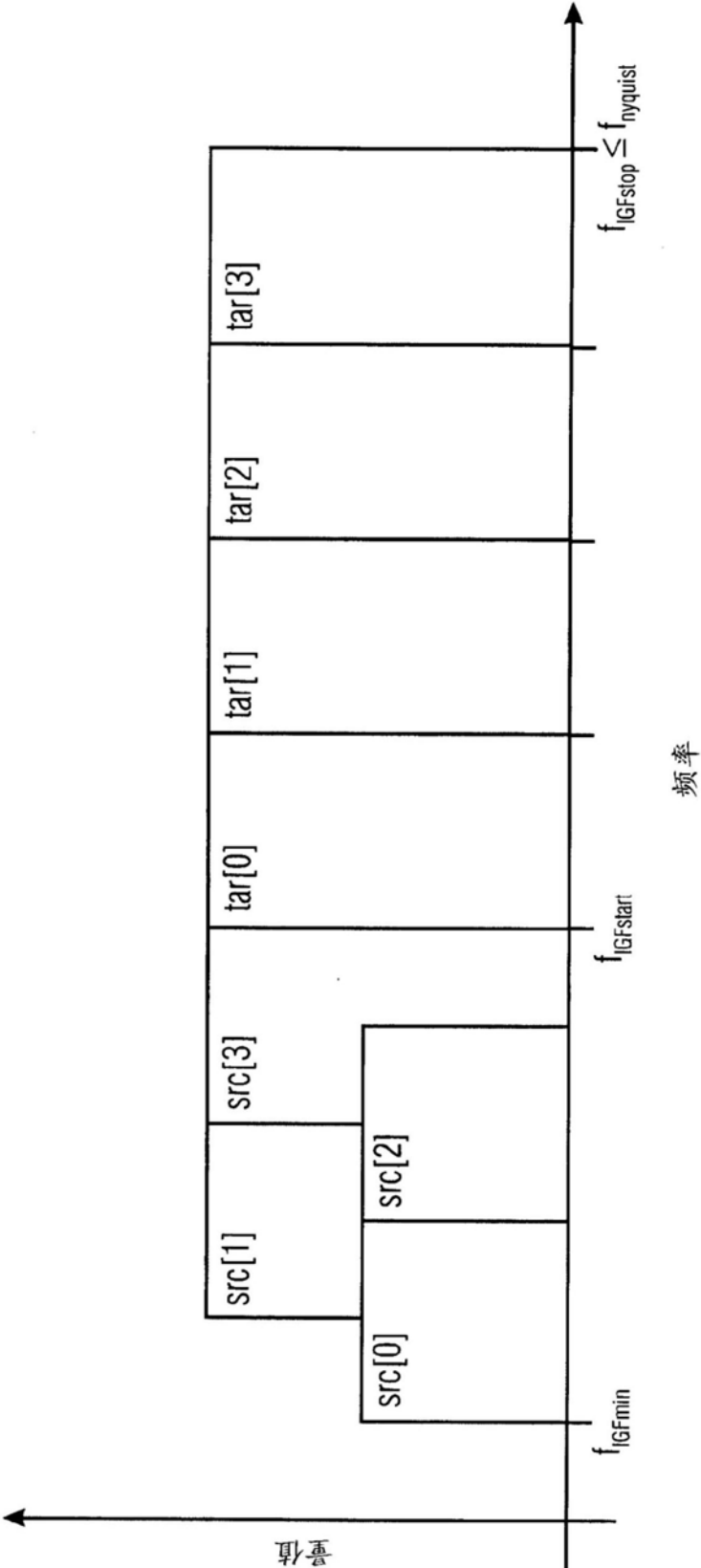


图12B

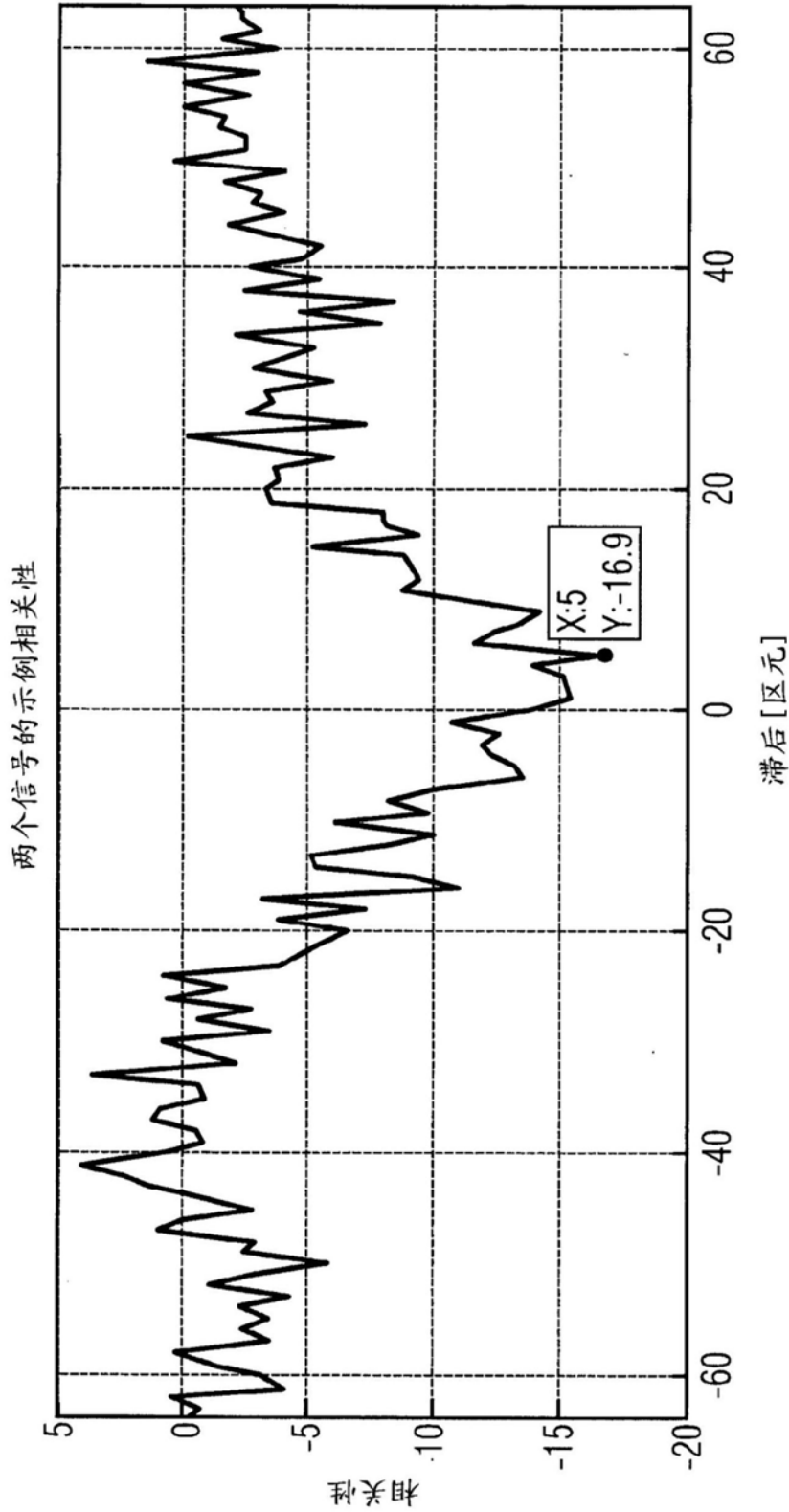


图12C

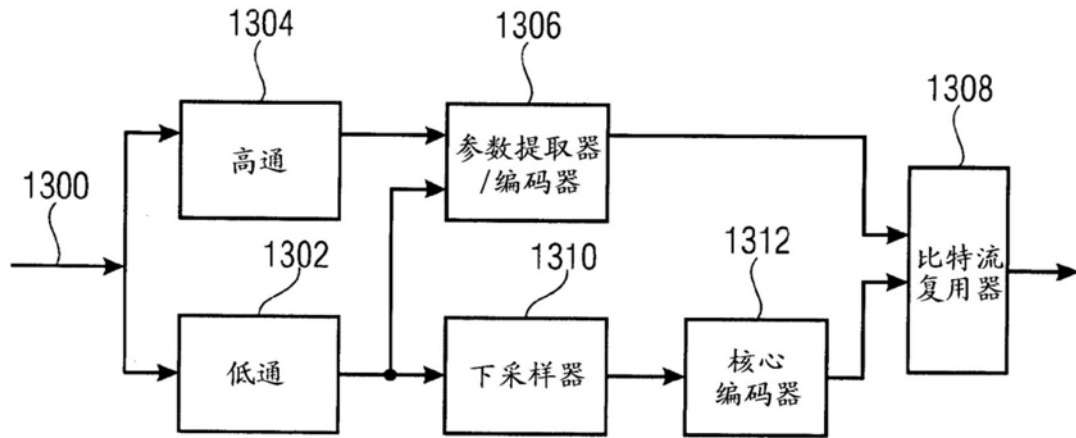


图13A (现有技术)

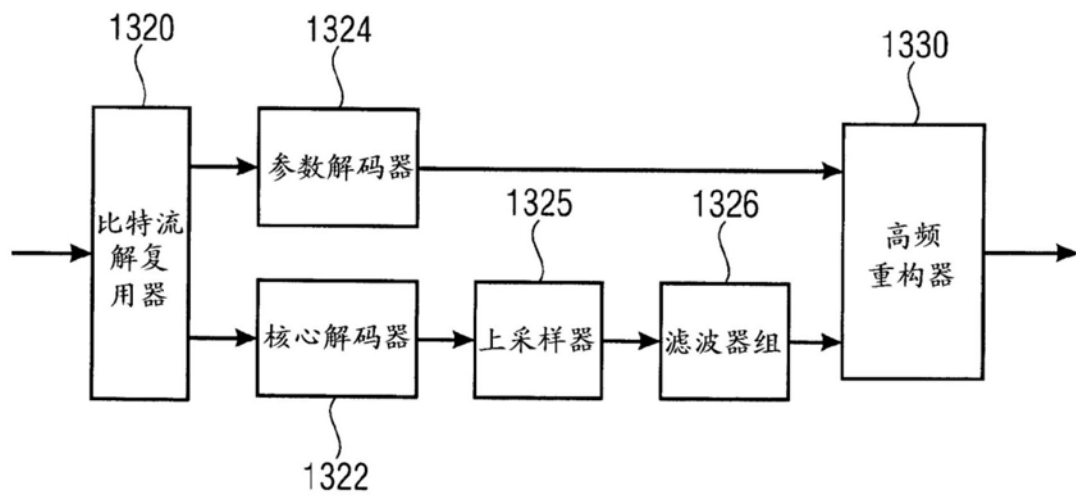


图13B (现有技术)