



(12) 发明专利

(10) 授权公告号 CN 103077722 B

(45) 授权公告日 2015. 07. 22

(21) 申请号 201210491312. 8

(22) 申请日 2009. 07. 06

(30) 优先权数据

61/079, 873 2008. 07. 11 US

(62) 分案原申请数据

200980135837. 4 2009. 07. 06

(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 斯特凡·拜尔 萨沙·迪施

拉尔夫·盖格尔 纪尧姆·福克斯

马克斯·诺伊恩多夫

杰拉尔德·舒勒 贝恩德·埃德勒

(74) 专利代理机构 中科专利商标代理有限责任

公司 11021

代理人 潘剑颖

(51) Int. Cl.

G10L 19/028(2013. 01)

G10L 19/002(2013. 01)

G10L 21/04(2013. 01)

(56) 对比文件

US 2007/0100607 A1, 2007. 05. 03, 全文.

WO 2006/113921 A1, 2006. 10. 26, 全文.

US 2003/0233234 A1, 2003. 12. 18, 全文.

CN 1440549 A, 2003. 09. 03, 全文.

R. J. Sluijter et al.. A TIME WARP
FOR SPEECH SIGNALS. 《1999 IEEE Workshop on
Speech Coding Proceedings》. 1999, 150-152.

审查员 康丹丹

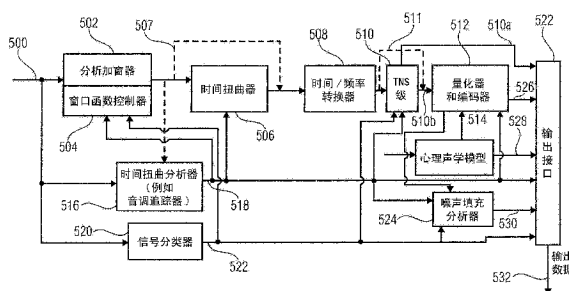
权利要求书3页 说明书25页 附图26页

(54) 发明名称

提供时间扭曲激活信号以及使用该时间扭曲
激活信号对音频信号编码

(57) 摘要

音频编码器包括窗口函数控制器 (504)、加
窗器 (502)、具有最终质量检查功能的时间扭曲
器 (506)、时间 / 频率转换器 (508)、TNS 级 (510)
或量化器编码器 (512), 由时间扭曲分析器 (516)
或信号分类器 (520) 获得的信号分析结果来控制
所述窗口函数控制器 (504)、所述时间扭曲器
(506)、所述 TNS 级 (510) 或附加的噪声填充分
析器 (524)。此外, 解码器使用取决于音频信号的
谐波或语音特性的经操纵的噪声填充估计来应用
噪声填充操作。



(编码器)

1. 一种用于基于音频信号的表示 (110 ;234e ;234k) 来提供时间扭曲激活信号 (112 ;232 ;234p) 的时间扭曲激活信号提供器 (100 ;230 ;234), 所述时间扭曲激活信号提供器包括:

能量压缩信息提供器 (120 ;234f ;234l ;325 ;370), 被配置为提供能量压缩信息 (122 ;234m ;234n ;326 ;374), 所述能量压缩信息描述所述音频信号的时间扭曲变换频谱表示 (222) 的能量压缩;以及

比较器 (130 ;234o), 被配置为将所述能量压缩信息 (122 ;234m ;234n ;326 ;374) 与参考值相比较, 以及被配置为取决于比较结果提供时间扭曲激活信号 (112 ;232 ;234p)。

2. 根据权利要求 1 所述的时间扭曲激活信号提供器 (100 ;230 ;234), 其中, 所述能量压缩信息提供器 (120 ;234f ;234l) 被配置为提供作为所述能量压缩信息 (122 ;234m ;234n) 的频谱平坦度度量, 所述频谱平坦度度量描述所述音频信号的时间扭曲变换频谱表示 (234e ;234k)。

3. 根据权利要求 2 所述的时间扭曲激活信号提供器 (100 ;230 ;234), 其中, 所述能量压缩信息提供器 (120 ;234f ;234l) 被配置为计算所述音频信号的时间扭曲变换功率频谱 (234e ;234k) 的几何平均与所述音频信号的时间扭曲变换功率频谱 (234e ;234k) 的算术平均的商, 以获得所述频谱平坦度度量。

4. 根据权利要求 1 所述的时间扭曲激活信号提供器 (100 ;230 ;234), 其中, 所述能量压缩信息提供器 (120 ;234f ;234l) 被配置为: 与所述时间扭曲变换频谱表示 (234e ;234k) 的较低频率部分相比时, 强调所述时间扭曲变换频谱表示 (234e ;234k) 的较高频率部分, 以获得所述能量压缩信息 (122 ;234m ;234n)。

5. 根据权利要求 1 所述的时间扭曲激活信号提供器 (100 ;230 ;234), 其中, 所述能量压缩信息提供器 (120 ;234m ;234n) 被配置为获得频谱平坦度的多个逐频带度量, 以及被配置为计算所述频谱平坦度的多个逐频带度量的平均值, 以获得所述能量压缩信息 (122, 234m ;234n)。

6. 根据权利要求 1 所述的时间扭曲激活信号提供器 (100 ;230 ;234), 其中, 所述能量压缩信息提供器 (120 ;234f ;234l ;325) 被配置为提供作为所述能量压缩信息 (122 ;234m ;234n) 的感知熵 (pe) 度量, 所述感知熵 (pe) 度量描述所述音频信号的时间扭曲变换频谱表示 (234e ;234k)。

7. 根据权利要求 6 所述的时间扭曲激活信号提供器 (100 ;230 ;234 ;235), 其中, 所述能量压缩信息提供器 (120 ;234f ;234l ;325) 被配置为基于扩缩因子频带的波形因子信息 (ffac(n)), 计算所述音频信号的时间扭曲变换频谱表示 (234e ;234k) 的一个或多个扩缩因子频带的非零线的估计数目 (n1), 以及被配置为将非零线的所述估计数目 (n1) 与在考察的扩缩因子频带的能量度量相乘, 来计算所考察的所述扩缩因子频带的感知熵 (326) 度量。

8. 根据权利要求 1 所述的时间扭曲激活信号提供器 (100 ;230 ;234), 其中, 所述能量压缩信息提供器 (120 ;234f ;234l ;370) 被配置为提供作为所述能量压缩信息的自相关度量 (374), 所述自相关度量 (374) 描述所述音频信号的时间扭曲时域表示 (234e ;234k) 的自相关。

9. 根据权利要求 8 所述的时间扭曲激活信号提供器 (100 ;230 ;234), 其中, 所述能量

压缩信息提供器 (120 ;234f ;234l ;370) 被配置为确定所述音频信号的时间扭曲时域表示 (234e ;234k) 的归一化自相关函数的绝对值之和, 以获得所述能量压缩信息。

10. 根据权利要求 1 所述的时间扭曲激活信号提供器 (100 ;230), 其中, 所述时间扭曲激活信号提供器包括参考值计算器, 所述参考值计算器被配置为基于所述音频信号的未扭曲频谱表示 (210)、或基于所述音频信号的未扭曲时域表示 (210), 来计算所述参考值; 以及

其中, 所述比较器被配置为使用描述所述音频信号的时间扭曲变换频谱表示的能量压缩的能量压缩信息 (122) 以及所述参考值, 来形成比值, 以及被配置为将所述比值与一个或多个阈值进行比较, 以获得作为比较结果的所述时间扭曲激活信号。

11. 根据权利要求 1 所述的时间扭曲激活信号提供器 (230 ;234), 其中, 所述时间扭曲激活信号提供器包括参考值计算器, 所述参考值计算器被配置为基于输入信号的时间扭曲表示 (210) 来计算所述参考值, 所述输入信号 (210) 的时间扭曲表示 (210) 是使用标准时间扭曲轮廓信息 (288) 来进行时间扭曲的; 以及

其中, 所述比较器被配置为使用描述所述音频信号的时间扭曲表示的能量压缩的所述能量压缩信息 (234e) 以及所述参考值, 来形成比值, 以及被配置为将所述比值与一个或多个阈值进行比较, 以获得作为比较结果的所述时间扭曲激活信号。

12. 一种用于对输入音频信号 (210) 编码以获得所述输入音频信号的编码表示 (212) 的音频信号编码器 (200), 所述音频信号编码器包括:

时间扭曲变换器 (220), 被配置为基于所述输入音频信号 (210), 使用时间扭曲轮廓来提供时间扭曲变换频谱表示 (222);

根据权利要求 1 所述的时间扭曲激活信号提供器 (100 ;230 ;234), 其中, 所述时间扭曲激活信号提供器被配置为接收所述输入音频信号 (210), 以及被配置为提供所述时间扭曲激活信号 (112 ;232 ;234p); 以及

控制器 (240), 被配置为取决于所述时间扭曲激活信号 (112 ;232 ;234p), 选择性地向所述时间扭曲变换器 (220) 提供描述非恒定时间扭曲轮廓部分的新发现的时间扭曲轮廓信息 (286), 或提供描述恒定时间扭曲轮廓部分的标准时间扭曲轮廓信息 (288), 以描述由所述时间扭曲变换器 (220) 所使用的时间扭曲轮廓。

13. 根据权利要求 12 所述的音频信号编码器, 其中, 所述音频信号编码器包括输出接口 (280), 所述输出接口 (280) 被配置为将所述时间扭曲变换频谱表示 (222) 包括在所述音频信号的编码表示 (212) 中, 以及

被配置为取决于所述时间扭曲激活信号 (232) 选择性地时间扭曲轮廓信息包括在所述音频信号的编码表示 (212) 中。

14. 一种基于音频信号提供时间扭曲激活信号的方法 (400), 所述方法包括:

提供 (410) 描述所述音频信号的时间扭曲变换频谱表示的能量压缩的能量压缩信息; 将所述能量压缩信息与参考值相比较 (420); 以及
取决于比较结果来提供 (430) 所述时间扭曲激活信号。

15. 一种用于对输入音频信号编码以获得所述输入音频信号的编码表示的方法 (450), 所述方法包括:

根据权利要求 14 所述的方法来提供 (470) 时间扭曲激活信号, 其中, 能量压缩信息描

述所述输入音频信号的时间扭曲变换频谱表示的能量压缩 ;以及

取决于所述时间扭曲激活信号,选择性地提供 (480) 所述输入音频信号的时间扭曲变换频谱表示的描述或所述输入音频信号的非时间扭曲变换频谱表示的描述,以将其包括在所述输入音频信号的编码表示中。

提供时间扭曲激活信号以及使用该时间扭曲激活信号对音频信号编码

[0001] 本申请是申请号为“200980135837.4”，申请日为2011年3月11日，发明名称为“提供时间扭曲激活信号以及使用该时间扭曲激活信号对音频信号编码”之申请的分案申请。

技术领域

[0002] 本发明涉及音频编码和解码，并且具体地针对具有谐波或语音内容的、可受到时间扭曲处理的音频信号的编码/解码。

背景技术

[0003] 在下文中，将给出对时间扭曲音频编码的领域的简要说明，该编码的概念可结合本发明的一些实施例一起应用。

[0004] 近年来，技术上的发展可将音频信号变换为频域表示，并且例如考虑到感知屏蔽阈值，可以对该频域表示进行有效地编码。如果发送编码频谱系数组的块长度很长，且如果仅相当小数目频谱系数远在该全局屏蔽阈值之上，同时很大数目的频谱系数在该全局屏蔽阈值附近或之下并可能因而被忽略（或以最小码长进行编码）时，该音频信号编码的概念特别有效。

[0005] 例如，基于余弦或基于正弦的调制的重叠变换通常由于它们的能量压缩性质而用于源编码的应用。即，对于具有恒定基本频率（音调）的谐波而言，它们将信号能量浓缩于小数目频谱分量（子频带）中，这导致了有效的信号表示。

[0006] 大体而言，应当将信号的（基本）音调理解为可与该信号频谱相区别的最低主频率。在普通语音模型中，该音调是由人类喉咙调制的激励信号的频率。如果仅一个单一基本频率存在，该频谱将极其简单，仅包括该基本频率及泛音。可以高效地对这种频谱编码。然而，对于具有变化音调的信号，对应于每个谐波分量的能量散布于若干变换系数上，因而导致编码效率的减少。

[0007] 为了克服编码效率的减少，在不均匀时间网络上对要编码的音频信号有效地重新采样。在随后的处理中，对通过不均匀重新采样所获得的采样位置就好像它们表示均匀时间网络上的值一样进行处理。该操作一般由短语“时间扭曲”来表示。可取决于该音调的时间变化来有利地选择采样时间，使得该音频信号的时间扭曲版本中的音调变化小于该音频信号的原始版本（时间扭曲前）中的音调变化。该音调变化也可用短语“时间扭曲轮廓”表示。在音频信号的时间扭曲之后，将该音频信号的时间扭曲版本转换为频域。该依赖于音调的时间扭曲具有如下效果：时间扭曲音频信号的频域表示一般地显示出将能量压缩成远远小于该原始音频信号（未被时间扭曲）的频域表示的频谱分量数目。

[0008] 在解码器侧，将该时间扭曲音频信号的频域表示转换回时域，使得该时间扭曲音频信号的时域表示在解码器侧可用。然而，在解码器侧重建时间扭曲音频信号的时域表示中，不包括该编码器侧输入音频信号的原始音调变化。因此，通过对时间扭曲音频信号的

解码器侧重建时域表示进行重新采样,来应用另一时间扭曲。为了在解码器处获得对编码器侧输入音频信号的良好重建,需要解码器侧时间扭曲至少近似编码器侧时间扭曲的反操作。为了获得恰当的时间扭曲,需要让允许调整解码器侧时间扭曲的信息在解码器处可用。

[0009] 因为一般要求将这种信息从音频信号编码器传输至音频信号解码器,需要将该发送所需的比特率保持为小,同时仍允许在解码器侧可靠重建所需的时间扭曲信息。

[0010] 鉴于上述讨论,需要创建一种概念,其允许有效应用音频编码器中时间扭曲概念的比特率。

发明内容

[0011] 本发明的目的是创建以下概念:基于在时间扭曲音频信号编码器或时间扭曲音频信号解码器中可用的信息,来增强由编码音频信号所提供的听觉印象。

[0012] 由根据权利要求 1 所述的用于基于音频信号的表示来提供时间扭曲激活信号的时间扭曲激活信号提供器、根据权利要求 12 所述的用于对输入音频信号编码的音频信号编码器、根据权利要求 14 所述的用于提供时间扭曲激活信号的方法、根据权利要求 15 所述的用于提供输入音频信号的编码表示的方法、或根据权利要求 16 所述的计算机程序来达成该目的。

[0013] 本发明的另一目的是提供一种增强的音频编码/解码方案,该方案提供较高的质量或较低的比特率。

[0014] 由根据权利要求 17、26、32、37 所述的音频编码器、根据权利要求 20 所述的音频解码器、根据权利要求 23、30、35 或 37 所述的音频编码方法、根据权利要求 24 所述的解码方法、或根据权利要求 25、31、36 或 43 所述的计算机程序来达成该目的。

[0015] 根据本发明的实施例与用于时间扭曲 MDCT 变换编码器的方法相关。一些实施例仅与编码器工具相关。然而,其它实施例还与解码器工具相关。

[0016] 本发明的实施例创建时间扭曲激活信号提供器,其用于基于音频信号的表示来提供时间扭曲激活信号。该时间扭曲激活信号提供器包括能量压缩信息提供器,被配置为提供能量压缩信息,该信息描述音频信号的时间扭曲变换频谱表示中的能量压缩。该时间扭曲激活信号提供器还包括比较器,该比较器被配置为将能量压缩信息与参考值相比较,且取决于比较结果来提供时间扭曲激活信号。

[0017] 该实施例基于如下发现:如果音频信号的时间扭曲变换频谱表示由于将能量集中于一个或多个频谱区域(或频谱线)而包括充分压缩的能量分布,则从编码音频信号的比特率减少的意义上来说,音频信号编码器中的时间扭曲功能性的使用一般带来增强。这是由于如下的事实:通过将模糊频谱(例如音频帧的模糊频谱)变换为具有一个或多个可辨别波峰的频谱,且因此变换为具有比原始(未时间扭曲)音频信号的频谱更高的能量压缩的频谱,则成功的时间扭曲带来减少比特率的效果。

[0018] 关于此问题,应理解音频信号帧(在该帧中音频信号的音调显著地变化)包括模糊频谱。音频信号的时间变化音调具有如下效果:在音频信号帧上执行的时域到频域的变换导致信号能量在频域,具体地在较高频域,上的模糊分布。因此,这种原始(未时间扭曲)音频信号的频谱表示包括低能量压缩,且一般在该频谱的较高频率部分不显示频谱波峰,或仅在频谱中较高频率部分显示相对小的频谱波峰。相对地,如果时间扭曲成功(就提供

该编码效率的增强而言),该原始音频信号的时间扭曲产生具有相对较高且清晰的波峰的频谱(具体地在该频谱的较高频率部分中)的时间扭曲音频信号。这是由于以下事实:将具有时间变化音调的音频信号变换为具有较小音调变化或甚至近似恒定音调的时间扭曲音频信号。因此,该时间扭曲音频信号的频谱表示(可以将其视为该音频信号的时间扭曲变换频谱表示)包括一个或多个清晰频谱波峰。换言之,通过成功的时间扭曲操作来减少该原始音频信号(具有在时间上变化的音调)频谱的模糊,使得该音频信号的时间扭曲变换频谱表示包括比原始音频信号的频谱更高的能量压缩。然而,时间扭曲在增强编码效率中并不总是成功。例如,如果输入音频信号包括大的噪声分量,或如果所提取的时间扭曲轮廓不精确,则时间扭曲不增强编码效率。

[0019] 鉴于该情况,由能量压缩信息提供器提供的能量压缩信息就减少比特率而言是判定该时间扭曲是否成功的有价值指示符。

[0020] 本发明的实施例创建时间扭曲激活信号提供器,用于基于音频信号的表示提供时间扭曲激活信号。该时间扭曲激活提供器包括两个时间扭曲表示提供器,所述两个时间扭曲标识提供器被配置为使用不同的时间扭曲轮廓信息来提供该相同音频信号的两个时间扭曲表示。因此,该时间扭曲表示提供器可以用相同的方式配置(在结构上或功能上),且使用相同音频信号但是不同的时间扭曲轮廓信息。该时间扭曲激活信号提供器还包括两个能量压缩信息提供器,所述两个能量压缩信息提供器被配置为基于第一时间扭曲表示提供第一能量压缩信息,且基于第二时间扭曲表示提供第二能量压缩信息。该能量压缩信息提供器可以用相同方式配置,但是使用不同的时间扭曲表示。此外,该时间扭曲激活信号提供器包括比较器,以将两个不同能量压缩信息进行比较,且提供取决于比较结果的时间扭曲激活信号。

[0021] 在优选实施例中,该能量压缩信息提供器被配置为提供作为能量压缩信息的频谱平坦度度量,该频谱平坦度度量描述该音频信号的时间扭曲变换频谱表示。已发现如果时间扭曲将输入音频信号变换为表示该输入音频信号的时间扭曲版本的较不平坦的时间扭曲频谱时,就减少比特率而言,时间扭曲是成功的。因此,频谱平坦度度量可以用于判定在不执行全频谱编码处理的情况下,应当激活还是停用时间扭曲。

[0022] 在优选实施例中,该能量压缩信息提供器被配置为计算该时间扭曲变换功率频谱的几何平均与该时间扭曲变换功率频谱的算术平均的商,以获得频谱平坦度度量。已发现该商是非常适于描述通过时间扭曲来获得的可能比特率节约的频谱平坦度度量。

[0023] 在另一优选实施例中,该能量压缩信息提供器被配置为当与时间扭曲变换频谱表示的较低频率部分相比时,强调时间扭曲变换频谱表示的较高频率部分,以获得该能量压缩信息。该概念基于如下发现:时间扭曲在较高频率范围上一般比在较低频率范围上具有更大的影响。因此,为了确定使用频谱平坦度度量的时间扭曲的有效性,主要评估该较高频率范围是恰当的。此外,典型的音频信号显示谐波内容(包括基本频率的谐波),其随频率的增加在强度上衰减。当与时间扭曲变换频谱表示的较低频率部分相比时,强调该时间扭曲变换频谱表示的较高频率部分也有助于补偿该频谱线随频率增加的这种典型衰减。总而言之,对频谱的较高频率部分的强调导致了能量压缩信息的可靠性增加,并因此允许更可靠地提供时间扭曲激活信号。

[0024] 在另一优选实施例中,能量压缩信息提供器被配置为提供频谱平坦度的多个逐频

带度量,且被配置为计算频谱平坦度的多个逐频带度量的平均值,以获得该能量压缩信息。已发现逐频带频谱平坦度度量的考虑导致了与时间扭曲是否有效减少编码音频信号比特率的特别可靠信息。首先,一般以逐频带方式来执行对时间扭曲变换频谱表示的编码,使得频谱平坦度的该逐频带度量的组合非常适于该编码,且因此以良好精确度表示可获得的比特率增强。此外,频谱平坦度度量的逐频带计算实质上消除了能量压缩信息对谐波分布的依赖性。例如,即使较高频带包括相对小的能量(小于较低频带的能量),该较高频带可能仍然在感知上是相关的。然而,如果不以逐频带方式来计算该频谱平坦度度量,则在该较高频带上的时间扭曲的积极影响(从该频谱线的模糊的减少的意义上说)可能仅因该较高频带上的能量小而被认为是小的。相对地,通过应用逐频带计算,可以用恰当的权重来考虑时间扭曲的积极影响,因为该逐频带频谱平坦度度量独立于各自频带中的绝对能量。

[0025] 在另一优选实施例中,该时间扭曲激活信号提供器包括参考值计算器,所述参考值计算器被配置为计算频谱平坦度度量,以获得该参考值,该度量描述音频信号的未时间扭曲的频谱表示。因此,可基于输入音频信号的未时间扭曲(或“未扭曲的”)版本的频谱平坦度与输入音频信号的时间扭曲版本的频谱平坦度的比较来提供该时间扭曲激活信号。

[0026] 在另一优选实施例中,该能量压缩信息提供器被配置为提供作为能量压缩信息的感知熵度量,该度量描述音频信号的时间扭曲变换频谱表示。此概念基于下述发现:时间扭曲变换频谱表示的感知熵是对编码该时间扭曲变换频谱所需要的比特数目(或比特率)的良好估计。因此,甚至由于如果使用时间扭曲,则必须对附加时间扭曲信息编码,该时间扭曲变换频谱表示的感知熵度量是是否可通过时间扭曲来预期比特率减少的良好度量。

[0027] 在另一优选实施例中,该能量压缩信息提供器被配置为提供作为能量压缩信息的自相关度量,该度量描述音频信号的时间扭曲表示的自相关。该概念基于如下发现:可以基于时间扭曲(或不均匀重新采样)的时域信号来测量(或至少估计)时间扭曲的效率(就减少比特率而言)。已发现如果时间扭曲时域信号包括由自相关度量反映的相对高度的周期性,则时间扭曲是有效率的。相对地,如果时间扭曲时域信号不包括显著的周期性,则可以推断该时间扭曲是无效率的。

[0028] 该发现基于如下事实:有效时间扭曲将变化频率(不包括周期性)的正弦信号的一部分变换为接近恒定频率(包括高度的周期性)的正弦信号的一部分。相对地,如果时间扭曲不能提供具有高度周期性的时域信号,那么可预期时间扭曲也不提供可证明其应用可行的显著比特率节约。

[0029] 在优选实施例中,该能量压缩信息提供器被配置为确定音频信号的时间扭曲表示的归一化自相关函数的绝对值之和(对多个延迟值),以获得该能量压缩信息。已发现在估计时间扭曲的效率上不要求对自相关峰值的计算复杂的确定。而是,已发现对(大)范围的自相关延迟值上的自相关的求和评估也产生非常可靠的结果。这是由于如下事实:时间扭曲实际上将变化频率的多个信号分量(例如,基本频率及其谐波)变换为周期性信号分量。因此,这种时间扭曲信号的自相关在多个自相关延迟值处显示波峰。因此,求和形式是从自相关提取能量压缩信息的计算上高效率的方式。

[0030] 在另一优选实施例中,该时间扭曲激活信号提供器包括参考值计算器,所述参考值计算器被配置为基于音频信号的未时间扭曲频谱表示,或基于音频信号的未时间扭曲时域表示,来计算参考值。在该情况中,比较器一般被配置为使用能量压缩信息及参考值形成

比值,该能量压缩信息描述音频信号的时间扭曲变换频谱的能量压缩。该比较器也被配置为将该比值与一个或多个阈值进行比较,以获得时间扭曲激活信号。已发现在未时间扭曲情况中的能量压缩信息与在时间扭曲情况中的能量压缩信息之间的比率允许产生计算上高效率但仍充分可靠的时间扭曲激活信号。

[0031] 本发明的另一优选实施例创建音频信号编码器,用于对输入音频信号编码,以获得该输入音频信号的编码表示。音频信号编码器包括时间扭曲变换器,被配置为基于输入音频信号,提供时间扭曲变换频谱表示。该音频信号编码器还包括如上所述的时间扭曲激活信号提供器。该时间扭曲激活信号提供器被配置为接收输入音频信号,且提供能量压缩信息,使得该能量压缩信息描述该输入音频信号的时间扭曲变换频谱表示中的能量压缩。该音频信号编码器还包括控制器,被配置为取决于时间扭曲激活信号,向时间扭曲变换器选择性地提供发现的非恒定(变化)时间扭曲轮廓部分或时间扭曲信息,或标准恒定(不变)时间扭曲轮廓部分或时间扭曲信息。这样,有可能选择性地接受或拒绝由该输入音频信号的编码音频信号表示推导出的发现的非恒定时间扭曲轮廓部分。

[0032] 该概念基于下述发现:将时间扭曲信息引入该输入音频信号的编码表示并不总是有效,因为要求相当可观数目的比特用于编码该时间扭曲信息。此外,已发现由时间扭曲激活信号提供器计算出的能量压缩信息是判定将该发现的变化(非恒定)时间扭曲估计部分还是标准(不变、恒定)时间扭曲轮廓提供给时间扭曲变换器是否有利的一种计算上高效率的度量。已注意到当该时间扭曲变换器包括重叠变换时,可在两个或更多随后的变换块的计算中使用发现的时间扭曲轮廓部分。具体地,已发现为了能做出时间扭曲是否允许比特率的节约的判定,并无必要使用新发现的变化时间扭曲轮廓部分对该输入音频信号的时间扭曲变换频谱表示版本进行完全编码,以及并无必要使用标准(不变)时间扭曲轮廓部分对该输入音频信号的时间扭曲变换频谱表示版本进行完全编码。而是,已发现对输入音频信号的时间扭曲变换频谱表示的能量压缩的评估形成了该判定的可靠基础。因此,可以将所需的比特率保持为小。

[0033] 在又一优选实施例中,该音频信号编码器包括输出接口,被配置为取决于时间扭曲激活信号,选择性地包括时间扭曲轮廓信息,该信息将发现的变化时间扭曲轮廓表示为该音频信号的编码表示。因此,可获得高效的音频信号编码,而不管该输入信号是否非常适合于时间扭曲。

[0034] 根据本发明的另一实施例创建一种基于音频信号来提供时间扭曲激活信号的方法。该方法实现时间扭曲激活信号提供器的功能,且可由本文中与时时间扭曲激活信号提供器相关描述的任何特征及功能来补充。

[0035] 根据本发明的另一实施例创建一种用于对输入音频信号编码,以获得输入音频信号的编码表示的方法。该方法可由本文中与时音频信号编码器相关描述的任何特征及功能来补充。

[0036] 根据本发明的另一实施例创建一种用于执行本文所述方法的计算机程序。

[0037] 根据本发明的第一方面,一种音频信号分析,有利地使用音频信号是具有谐波特性还是语音特性,用于控制编码器侧和/或解码器侧的噪声填充处理。在使用时间扭曲功能的系统中易于获得该音频信号分析,因为时间扭曲功能一般包括音调追踪器和/或信号分类器,用于区分语音与音乐,和/或区分有发音语音与无发音语音。因为该信息在这种上

下文中可用而不需任何此外的成本,因此可用的信息有利地用于控制该噪声填充特征,使得尤其对于语音信号,可减少谐波线之间的噪声填充,或具体地对于语音信号,甚至消除谐波线之间的噪声填充。甚至在获得强谐波内容但是语音检测器没有直接检测到语音的情况下,噪声填充的减少仍然将导致更高的感知质量。虽然该特征在无论如何也执行谐波 / 语音分析的系统中特别有用,且因此该信息可用且不需任何附加成本,甚至当必须将特定信号分析器插入该系统中时,对基于信号具有谐波还是语音特性的信号分析的噪声填充方案的控制也是附加有用的,因为增强质量而比特率没有增加,或换言之,比特率减少而质量没有损失,因此当减少可从编码器发送至解码器的噪声填充级别本身时,减少了用于对该噪声填充级别编码所需的比特。

[0038] 在本发明另一方面中,信号分析结果,即信号是谐波信号还是语音信号,用于控制音频编码器的窗口函数处理。已发现在语音信号或谐波信号开始的情况下,简单编码器将从长窗口切换至短窗口的可能性是很高的。然而这些短窗口具有对应地减少的频谱解析度,另一方面,该频率解析度将减少强谐波信号的编码增益,且因此增加对这种信号部分编码所需的比特数目。鉴于此,当检测到语音或谐波信号开始时,在本方面中定义的本发明使用比短窗口更长的窗口。备选地,选择具有与该长窗口大致相似长度的但具有更短重叠的窗口,以有效地减少前回声。大体上,音频信号的时帧具有谐波还是语音特性的信号特性用于选择针对该时帧的窗口函数。

[0039] 根据本发明的另一方面,基于底层信号是基于时间扭曲操作还是在线性域中来控制 TNS(时域噪声修整)工具。一般地,已通过时间扭曲操作来处理的信号将具有强谐波内容。否则,与时间扭曲级相关联的音调追踪器将不会输出有效音调轮廓,且在缺少这种有效音调轮廓时,对与音频信号的该时帧将停用时间扭曲功能。然而,谐波信号将一般不适于经受 TNS 处理。当由 TNS 级处理的信号具有相当平坦的频谱时, TNS 处理特别有用且产生比特率 / 质量上的重要增益。然而,当该信号的外观是音调的 (tonal), 即非平坦的,如同在具有谐波内容或有发音内容的频谱的情况下,则将减少由 TNS 工具提供的质量 / 比特率上的增益。因此,不使用该 TNS 工具的发明性修改,时间扭曲部分一般不由 TNS 处理,但是会在不使用 TNS 滤波的情况下来处理。另一方面, TNS 的噪声修整特征仍然提供增强的质量,特别是在信号在振幅 / 功率上变化的情况下。在谐波信号或语音信号的开始存在,以及实施了块切换特征使得维持长窗口或者至少长于短窗口的窗口、而非该起始的情况下,该帧的时域噪声修整特征的激活将导致语音开始周围的噪声的浓缩,这有效地减少可能由于在随后的编码器处理中发生的帧量化而在语音开始之前发生的前回声。

[0040] 根据本发明的另一方面,由音频编码设备中的量化器 / 熵编码器来处理可变数目的线,以计入可变带宽,通过执行具有可变时间扭曲特性 / 扭曲轮廓的时间扭曲操作来引入该可变带宽。当该时间扭曲操作导致增加了时间扭曲帧中包括的帧时间(以线性)时,减少了单一频率线的带宽,且,对于恒定总带宽,在未时间扭曲情况下将增加要处理的频率线数目。另一方面,当时间扭曲操作导致在该时间扭曲域中音频信号的实际时间相对于在线性域中的音频信号块长度减少时,增加了单一频率线的频率带宽,且因此在未时间扭曲情况下,必须减少由源编码器处理的线数目,以具有减少的带宽变化或最好没有带宽变化。

附图说明

- [0041] 随后通过附图来描述优选实施例,其中;
- [0042] 图 1 示出了根据本发明的实施例的时间扭曲激活信号提供器的示意框图;
- [0043] 图 2a 示出了根据本发明的实施例的音频信号编码器的示意框图;
- [0044] 图 2b 示出了根据本发明的实施例的时间扭曲激活信号提供器的另一示意框图;
- [0045] 图 3a 示出了音频信号的未时间扭曲版本的频谱的图形表示;
- [0046] 图 3b 示出了音频信号的时间扭曲版本的频谱的图形表示;
- [0047] 图 3c 示出了针对不同频带的频谱平坦度度量的个别计算的图形表示;
- [0048] 图 3d 示出了仅考虑频谱的较高频带部分的频谱平坦度度量的计算的图形表示;
- [0049] 图 3e 示出了使用频谱表示的频谱平坦度度量的计算的图形表示,在该频谱表示中,相对于较低频率部分强调了较高频率部分;
- [0050] 图 3f 示出了根据本发明的另一实施例的能量压缩信息提供器的示意框图;
- [0051] 图 3g 示出了在时域中具有时间上可变音调的音频信号的图形表示;
- [0052] 图 3h 示出了图 3g 的音频信号的时间扭曲(不均匀重新采样的)版本的图形表示;
- [0053] 图 3i 示出了根据图 3g 的音频信号的自相关函数的图形表示;
- [0054] 图 3j 示出了根据图 3h 的音频信号的自相关函数的图形表示;
- [0055] 图 3k 示出了根据本发明另一实施例的能量压缩信息提供器的示意框图;
- [0056] 图 4a 示出了用于基于音频信号来提供时间扭曲激活信号的方法的流程图;
- [0057] 图 4b 示出了根据本发明的实施例的用于对输入音频信号编码,以获得该输入音频信号的编码表示的方法的流程图;
- [0058] 图 5a 示出了具有创造性方面的音频编码器的优选实施例;
- [0059] 图 5b 示出了具有创造性方面的音频解码器的优选实施例;
- [0060] 图 6a 示出了本发明的噪声填充方面的优选实施例;
- [0061] 图 6b 示出了定义由噪声填充级别操纵器所执行的控制操作的表格;
- [0062] 图 7a 示出了根据本发明的用于执行基于时间扭曲的块切换的优选实施例;
- [0063] 图 7b 示出了影响窗口函数的备选实施例;
- [0064] 图 7c 示出了用于基于时间扭曲信息来说明窗口函数的另一备选实施例;
- [0065] 图 7d 示出了在有发音启动处的正常 AAC 行为的窗口序列;
- [0066] 图 7e 示出了根据本发明的优选实施例获得的备选窗口序列;
- [0067] 图 8a 示出了 TNS(时域噪声整修)工具的基于时间扭曲的控制的优选实施例;
- [0068] 图 8b 示出了定义图 8a 中阈值控制信号产生器中所执行的控制步骤的表格;
- [0069] 图 9a-9e 示出了不同的时间扭曲特性以及在解码器侧时间扭曲操作之后发生的对音频信号的带宽上的对应影响;
- [0070] 图 10a 示出了用于控制编码处理器中的线的数目的控制器的优选实施例;
- [0071] 图 10b 示出了针对采样率要丢弃/添加的线的数目之间的依赖性;
- [0072] 图 11 示出了线性时间尺度与扭曲时间尺度之间的比较;
- [0073] 图 12a 示出了在带宽扩展的上下文中的实施;以及
- [0074] 图 12b 示出了表,该表示出了在时间扭曲域中的本地采样率与频谱系数的控制之间的依赖性。

具体实施方式

[0075] 图 1 示出了根据本发明的实施例的时间扭曲激活信号提供器的示意框图。该时间扭曲激活信号提供器 100 被配置为接收音频信号的表示 110, 且基于该表示 110 提供时间扭曲激活信号 112。时间扭曲激活信号提供器 100 包括能量压缩信息提供器 120, 被配置为提供能量压缩信息 122, 该信息 122 描述该音频信号的时间扭曲变换频谱表示的能量的压缩。时间扭曲激活信号提供器 100 还包括比较器 130, 被配置为将能量压缩信息 122 与参考值 132 作比较, 以取决于该比较的结果提供时间扭曲激活信号 112。

[0076] 如上所述, 已发现能量压缩信息是允许对时间扭曲是否带来比特节约的计算上高效率估计的有价值信息。已发现比特节约的存在性与该时间扭曲是否导致能量压缩的问题密切相关。

[0077] 图 2a 示出了根据本发明的实施例的音频信号编码器 200 的示意框图。音频信号编码器 200 被配置为接收输入音频信号 210 (也以 $a(t)$ 标明), 且基于该输入音频信号 210 提供该输入音频信号 210 的编码表示 212。音频信号编码器 200 包括时间扭曲变换器 220, 被配置为接收输入音频信号 210 (可在时域中表示该信号), 且基于输入音频信号 210 提供该输入音频信号 210 的时间扭曲变换频谱表示 222。音频信号编码器 200 还包括时间扭曲分析器 284, 被配置为分析输入音频信号 210, 且基于该输入音频信号 210, 提供时间扭曲轮廓信息 286 (例如绝对或相对时间扭曲轮廓信息)。

[0078] 音频信号编码器 200 还包括切换机制, 例如具有受控开关 240 的形式的切换机制, 以判定是发现的时间扭曲轮廓信息 286 还是标准时间扭曲轮廓信息 288 用于进一步的处理。因此, 该切换机制 240 被配置为取决于时间扭曲激活信息, 选择性地将发现的时间扭曲轮廓信息 286 或标准时间扭曲轮廓信息 288 作为新时间扭曲轮廓信息 242 提供给例如时间扭曲变换器 220 用于进一步的处理。应注意, 时间扭曲变换器 220 可例如针对音频帧的时间扭曲来使用新时间扭曲轮廓信息 242 (例如新时间扭曲轮廓部分), 且此外使用之前获得的时间扭曲信息 (例如一个或多个之前获得的时间扭曲轮廓部分)。该可选的频谱后处理可包括例如时域噪声整修和 / 或噪声填充分析。音频信号编码器 200 还包括量化器 / 编码器 260, 被配置为接收频谱表示 222 (可选地由频谱后处理 250 来处理), 且量化及编码该变换频谱表示 222。为此, 量化器 / 编码器 260 可与感知模型 270 耦合, 且从感知模型 270 接收感知关联信息 272, 以考虑感知屏蔽且根据人类感知以不同的频率槽来调整量化精确度。音频信号编码器 200 还包括输出接口 280, 被配置为基于由量化器 / 编码器 260 所提供的已量化且编码的频谱表示 262, 提供该音频信号的编码表示 212。

[0079] 音频信号编码器 200 还包括时间扭曲激活信号提供器 230, 被配置为提供时间扭曲激活信号 232。时间扭曲激活信号 232 例如可用于控制切换机制 240, 以判定新发现时间扭曲轮廓信息 286 还是标准时间扭曲轮廓信息 288 用于进一步的处理步骤中 (例如由时间扭曲变换器 220)。此外, 时间扭曲激活信息 232 可用于开关 280 中, 以判定输入音频信号 210 的编码表示 212 是否包括已选择的新时间扭曲轮廓信息 242 (从新发现时间扭曲轮廓信息 286 及标准时间扭曲轮廓信息中选择的)。一般地, 如果已选择时间扭曲轮廓信息描述非恒定 (变化) 时间扭曲轮廓, 则时间扭曲轮廓信息仅被包括在该音频信号的编码表示 212 中。同样, 编码表示 212 可包括时间扭曲激活信息 232 其本身, 例如具有指示该时间扭

曲激活或停用的一比特旗标的形式。

[0080] 为了利于理解,应注意时间扭曲变换器 220 一般包括分析加窗器 220a、重新采样器或“时间扭曲器”220b 及频谱域变换器(或时间/频率转换器)220c。然而,视实施而定,可将时间扭曲器 220b 放置于在信号处理方向上的分析加窗器 220a 之前。然而,在一些实施例中可将时间扭曲及时域到频谱域变换结合在单一单元中。

[0081] 在下文中,将描述关于时间扭曲激活信号提供器 230 的操作的细节。应注意时间扭曲激活信号提供器 230 可等效于时间扭曲激活信号提供器 100。

[0082] 时间扭曲激活信号提供器 230 优选地被配置为接收时域音频信号表示 210(也以 $a(t)$ 标明)、新发现时间扭曲轮廓信息 286,及标准时间扭曲轮廓信息 288。时间扭曲激活信号提供器 230 也被配置为使用时域音频信号 210、新发现时间扭曲轮廓信息 286 及标准时间扭曲轮廓信息 288,来获得描述由于新发现时间扭曲轮廓信息 286 而产生的能量压缩的能量压缩信息,且基于该能量压缩信息来提供时间扭曲激活信号 232。

[0083] 图 2b 示出了根据本发明的实施例的时间扭曲激活信号提供器 234 的示意框图。时间扭曲激活信号提供器 234 在一些实施例中可发挥时间扭曲激活信号提供器 230 的作用。时间扭曲激活信号提供器 234 被配置为接收输入音频信号 210,及两个时间扭曲轮廓信息 286 与 288,且基于它们来提供时间扭曲激活信号 234p。时间扭曲激活信号 234p 可发挥时间扭曲激活信号 232 的作用。时间扭曲激活信号提供器包括两个相同的时间扭曲表示提供器 234a、234g,被配置为分别接收输入音频信号 210 及时间扭曲轮廓信息 286 与 288,且基于它们分别提供两个时间扭曲表示 234e 及 234k。时间扭曲激活信号提供器 234 还包括两个相同的能量压缩信息提供器 234f 及 234l,被配置为分别接收时间扭曲表示 234e 及 234k,且基于它们分别提供能量压缩信息 234m 及 234n。时间扭曲激活信号提供器还包括比较器 234o,被配置为接收能量压缩信息 234m 及 234n,且基于它们提供时间扭曲激活信号 234p。

[0084] 为了利于理解,应注意时间扭曲表示提供器 234a 与 234g 一般包括(可选)相同的分析加窗器 234b 及 234h、相同的重新采样器或时间扭曲器 234c 及 234i,及(可选)相同的频谱域变换器 234d 及 234j。

[0085] 在下文中,将讨论用于获得能量压缩信息不同概念。事先将做介绍以说明典型音频信号上的时间扭曲效果。

[0086] 在下文中,将参考图 3a 及 3b 来描述音频信号上时间扭曲的效果。图 3a 示出了音频信号的频谱的图形表示。横坐标 301 描述频率,纵坐标 302 描述该音频信号的强度。曲线 303 描述了与频率 f 相关的非时间扭曲音频信号的强度。

[0087] 图 3b 示出了图 3a 中表示的音频信号的时间扭曲版本的频谱的图形表示。同样,横坐标 306 描述频率,纵坐标 307 描述该音频信号的扭曲版本的强度。曲线 308 描述该音频信号的时间扭曲版本的强度对频率。从图 3a 与 3b 的图形表示的比较可看出,该音频信号的未时间扭曲(“未扭曲”)版本包括模糊频谱,具体地在较高频域中。相对地,该输入音频信号的时间扭曲版本包括具有清晰可区分的频谱波峰的频谱,甚至在较高频域中。此外,甚至可在该输入音频信号的时间扭曲版本的较低频谱域中看到频谱波峰的中等锐化。

[0088] 应注意图 3b 中所示的输入音频信号的时间扭曲版本的频谱可由例如量化器/编码器 260 以比图 3a 所示的未扭曲输入音频信号的频谱更低的比特率来量化及编码。这是

由于如下事实：模糊频谱一般包括很大数目的感知相关频谱系数（即相对很小数目的被量化为零或被量化为很小值的频谱系数），同时如图 3 所示的“不那么平坦的”频谱一般包括较大数目被量化为零或被量化为很小值的频谱系数。可以用比被量化为较高值的频谱系数更少的比特来对被量化为零或被量化为很小值的频谱系数进行编码，使得可使用比图 3a 的频谱更少的比特对图 3b 的频谱编码。

[0089] 然而，还应注意时间扭曲的使用不总是导致时间扭曲信号的编码效率的显著增强。因此，在一些情况中，对时间扭曲信息（例如时间扭曲轮廓）编码所需的价格（在比特率的意义上）可能超出用于对时间扭曲变换频谱编码的节约（在比特率的意义上）（当与编码非时间扭曲变换频谱相比较时）。在此情况中，优选地使用标准（不变）时间扭曲轮廓提供该音频信号的编码表示，以控制该时间扭曲变换。因此，可忽略任何时间扭曲信息（即时间扭曲轮廓信息）的发送（除指示该时间扭曲的停用的旗标之外），从而保持该比特率很低。

[0090] 在下文中，将参考图 3c-3k 来描述用于对时间扭曲激活信号 112、232、234p 的可靠且计算上高效率的计算的不同概念。然而，在此之前，将简短概括该创造性概念的背景。

[0091] 基本假定是对具有变化音调的谐波信号应用时间扭曲使得该音调恒定，且使该音调恒定增强了通过随后的时间频率变换所获得的频谱的编码，因为仅有限数目的重要的线保留（参见图 3b），而不是若干频谱容量上不同谐波的模糊（参见图 3a）。然而，即使当检测到音调变化时，可忽略（例如，如果在谐波信号下有强噪声，或如果该变化太小以至较高谐波的模糊没有问题）编码增益上的增强（即所节约的比特的数量），或编码增益上的增强可少于需要将时间扭曲轮廓传输至解码器的比特的数量，或可简单地是错的。在这些情况中，优选地拒绝由时间扭曲轮廓编码器产生的变化时间扭曲轮廓（例如 286），而相反使用有效的一比特信令，以信号方式发送标准（不变）时间扭曲轮廓。

[0092] 本发明的范围包括创建一种判定已获得的时间扭曲轮廓部分是否提供足够的编码增益（例如足以补偿时间扭曲轮廓编码所需的开销的编码增益）的方法。

[0093] 如上所述，时间扭曲的最重要的方面是较少数目线的频谱能量压缩（参见图 3a 及 3b）。它们示出了能量压缩还对应于“不那么平坦的”的频谱（参见图 3a 及 3b），因为增加了该频谱的波峰与波谷之间的差。将该能量浓缩于较少的线处，所述较少的线在具有比之前更少能量的线之间。

[0094] 图 3a 与 3b 示出了具有强谐波及音调变化的帧的未扭曲频谱（图 3a）与同一帧的时间扭曲版本的频谱（图 3b）的示意性示例。

[0095] 鉴于该情况，已发现将频谱平坦度度量用作该时间扭曲效率的可能的度量是有利的。

[0096] 可例如通过功率频谱的几何平均除以功率频谱的算术平均来计算该频谱平坦度。例如，可根据如下公式来计算该频谱平坦度（也以“平坦度”简短地标明）：

[0097]

$$\text{平坦度} = \frac{\sqrt[N]{\prod_{n=0}^{N-1} x(n)}}{\left(\frac{\sum_{n=0}^{N-1} x(n)}{N} \right)}$$

[0098] 在上式中，x(n) 表示容量号码 n 的大小。此外，在上式中，N 表示该频谱平坦度度

量的计算所考虑到的频谱容量的总数目。

[0099] 在本发明的实施例中,可使用时间扭曲变换频谱表示 234e、234k 来执行作为能量压缩信息的“平坦度”的上述计算,使得可以保持如下关系:

$$[0100] \quad x(n) = |X|_{tw}(n)$$

[0101] 在该情况中, N 可以等于由频谱域变换器 234d、234j 提供的频谱线的数目, $|X|_{tw}(n)$ 是时间扭曲变换频谱表示 234e、234k。

[0102] 尽管该频谱度量是用于提供该时间扭曲激活信号的有用的量,类似于信号对噪声比 (SNR) 度量,该频谱平坦度度量的一个缺点是如果应用于整个频谱,则其强调具有较高能量的部分。通常,谐波频谱具有特定的频谱倾斜,意指大部分能量浓缩于头几个部分音调,然后随频率的增加而减少,导致该度量中较高部分的代表性不足。这在一些实施例中是不想要的,由于需要增强这些较高部分的质量,因为它们变得最模糊(参见图 3a)。在下文中,将讨论该频谱平坦度度量的关联性的增强的若干可选概念。

[0103] 在根据本发明的实施例中,选择一种与所谓的“分段式 SNR”度量相似的方法,导致逐频带频谱平坦度度量。在一定数目的频带中(例如分别地)执行该频谱平坦度度量的计算,且采用主要部分(或平均)。不同频带可具有相等的带宽。然而,优选地,这些带宽将遵循感知尺度,如关键频带,或对应于例如所谓的“高级音频编码”(也称为 AAC)的扩缩因子频带。

[0104] 将在下文中参考图 3c 来简短解释上述概念,图 3c 示出了针对不同频带的频谱平坦度度量的单独计算的图形表示。如图所示,可将该频谱分为不同的频带 311、312、313,它们可具有相等的带宽或可具有不同的带宽。例如,针对第一频带 311,可使用例如上文给出的“平坦度”公式来计算第一频谱平坦度度量。在该计算中,可以考虑第一频带的频率槽(游动变量 n 可采用第一频带的频率槽的频率槽索引),且可以考虑该第一频带 311 的宽度(可变 N 可采用以第一频带的频率槽为单位的宽度)。因此,获得针对第一频带 311 的平坦度度量。相似地,可考虑到第二频带 312 的频率槽及第二频带的宽度来计算针对第二频带 312 的平坦度度量。此外,可以用相同方法来计算附加频带如第三频带 313 的平坦度度量。

[0105] 随后,可以计算针对不同频带 311、312、313 的平坦度度量的平均值,且该平均值可用作能量压缩信息。

[0106] 另一方法(用于该时间扭曲激活信号的导出的增强)是将该频谱平坦度度量仅应用于特定频率。图 3d 示出了这种方法。如图所示,针对该频谱平坦度度量的计算,仅考虑在频谱的高频部分 316 中的频率槽。针对该频谱平坦度度量的计算忽略该频谱的低频部分。针对该频谱平坦度度量的计算,可以逐频带的考虑高频部分 316。备选地,针对该频谱平坦度度量的计算,可以作为整体地考虑全部高频部分 316。

[0107] 综上所述,可以将频谱平坦度的减少(由时间扭曲的应用引起的)视为该时间扭曲的效果的第一度量。

[0108] 例如,时间扭曲激活信号提供者 100、230、234(或其比较器 130、234o)可使用标准时间扭曲轮廓信息,将时间扭曲变换频谱表示 234e 的频谱平坦度度量与时间扭曲变换频谱表示 234k 的频谱平坦度度量进行比较,且基于所述比较来判定该时间扭曲激活信号是有效还是无效的。例如,当与没有时间扭曲的情况相比时,如果该时间扭曲导致频谱平坦度度量的充分减少,则通过时间扭曲激活信号的恰当设置来激活该时间扭曲。

[0109] 除上述方法以外,针对该频谱平坦度的计算,可相对于低频部分来强调该频谱的高频部分(例如通过恰当的扩缩)。图 3c 示出了时间扭曲变换频谱的图形表示,在该时间扭曲变换频谱中,相对于低频部分强调了高频部分。因此,补偿了该频谱中的高频部分的代表性不足。因此如图 3e 所示,可在完成扩缩的、其中相对于低频率槽强调了高频率槽的频谱上计算平坦度量。

[0110] 就比特节约而言,编码效率的典型度量将是感知熵,可以用一种如以下文献所述的方式来定义感知熵,使得其与对特定频谱进行编码所需的比特实际数目很好的联系起来:3GPP TS 26.403V7.0.0:3rdGeneration Partnership Project; Technical Specification Group Services and System Aspects; General audio codec audio processing functions; Enhanced AAC part; Section 5.6.1.1.3 Relation between bit demand and perceptual entropy。所以,该感知熵的减少是时间扭曲的效率的另一度量。

[0111] 图 3f 示出了能量压缩信息提供器 325,可取代能量压缩信息提供器 120、234f、234l,且可用在时间扭曲激活信号提供器 100、290、234 中。能量压缩信息提供器 325 被配置为接收该音频信号的表示,例如,以时间扭曲变换频谱表示 234e、234k 的形式,也以 $|X|_{tw}$ 标明。能量压缩信息提供器 325 还被配置为提供感知熵信息 326,可取代能量压缩信息 122、234m、234n。

[0112] 能量压缩信息提供器 325 包括波形因子计算器 327,被配置为接收时间扭曲变换频谱表示 234e、234k,且基于它们来提供波形因子信息 328,该波形因子信息 328 可与频带相关联。能量压缩信息提供器 325 还包括频带能量计算器 329,被配置为基于时间扭曲频谱表示 234e、234k 来计算频带能量信息 $en(n)$ (330)。能量压缩信息提供器 325 还包括线数目估计器 331,被配置为对具有索引 n 的频带提供线的估计数目的信息 $n1$ (332)。此外,能量压缩信息提供器 325 包括感知熵计算器 333,被配置为基于频带能量信息 330 及线的估计数目的信息 332,计算感知熵信息 326。例如,波形因子计算器 327 可被配置为根据下述公式来计算波形因子:

$$[0113] \quad \text{ffac}(n) = \sum_{k=k\text{Offset}(n)}^{k\text{Offset}(n+1)-1} \sqrt{|X(k)|} \quad (1)$$

[0114] 在上述公式中, $\text{ffac}(n)$ 表示具有频带索引 n 的频带的波形因子。 k 表示游动变量,在扩缩因子频带(或频带) n 的频谱容量索引上游动。 $X(k)$ 表示具有频谱容量索引(或频率槽索引) k 的频谱容量(或频率槽)的频谱值(例如,能量值或数量值)。

[0115] 线数目估计器可被配置为根据如下公式估计非零线的数目,由 $n1$ 表示:

$$[0116] \quad n1 = \frac{\text{ffac}(n)}{\left(\frac{\text{en}(n)}{k\text{Offset}(n+1) - k\text{Offset}(n)} \right)^{0.25}} \quad (2)$$

[0117] 在上述公式中, $\text{en}(n)$ 表示具有索引 n 的频带或扩缩因子频带的能量。 $k\text{Offset}(n+1) - k\text{Offset}(n)$ 表示以频谱容量为单位的具有索引 n 的频带或扩缩因子频带的宽度。

[0118] 此外,感知熵计算器 332 可被配置为根据如下公式计算感知熵信息 sfbPe :

$$[0119] \quad sfbPe = nl \cdot \begin{cases} \log_2\left(\frac{en}{thr}\right) & \text{for } \log_2\left(\frac{en}{thr}\right) \geq c1 \\ (c2 + c3 \cdot \log_2\left(\frac{en}{thr}\right)) & \text{for } \log_2\left(\frac{en}{thr}\right) < c1 \end{cases} \quad (3)$$

[0120] 在上文中,下述关系将被保持:

$$[0121] \quad c1 = \log_2(8) \quad c2 = \log_2(2.5) \quad c3 = 1 - c2/c1 \quad (4)$$

[0122] 可将总的感知熵 pe 计算为多个频带或扩缩因子频带的感知熵之和。

[0123] 如上所述,感知熵信息 326 可用作能量压缩信息。

[0124] 对于关于感知熵的计算的其他细节,参考国际标准“3GPP TS26.403V7.0.0(2006-06)”的第 5.6.1.1.3 节。

[0125] 在下文中,将描述针对时域中的能量压缩信息的计算的概念。

[0126] 再看 TW-MDCT(时间扭曲改良型离散余弦变换)是以一种方式改变该信号,以具有一个块中的恒定或几乎恒定音调的基本观念。如果达成恒定音调,这意味着一个处理块的自相关的最大值增加。由于发现对于时间扭曲及未时间扭曲情况的相对应的自相关中的最大值是不平凡的,可以将归一化自相关的绝对值之和用作针对该增强的度量。该和的增加对应于能量压缩的增加。

[0127] 将在下文中参考图 3g、3h、3i、3j 及 3k 来更详细地解释该概念。

[0128] 图 3g 示出了时域中未时间扭曲信号的图形表示。横坐标 350 描述时间,纵坐标 351 描述未时间扭曲时间信号的级别 $a(t)$ 。曲线 352 描述未时间扭曲时间信号的时间上的演进。假定如图 3g 所示,由曲线 352 描述的未时间扭曲时间信号的频率随时间增加。

[0129] 图 3h 示出了图 3g 的时间信号的时间扭曲版本的图形表示。横坐标 355 示出了扭曲时间(例如以归一化的形式),纵坐标 356 示出了信号 $a(t)$ 的时间扭曲版本 $a(t_w)$ 的级别。如图 3h 所示,未时间扭曲时间信号 $a(t)$ 的时间扭曲版本 $a(t_w)$ 包括(至少近似地)在扭曲时域中的时间上恒定的频率。

[0130] 换言之,图 3h 示出了如下事实:将时间上变化的频率的时间信号通过恰当的时间扭曲操作变换为时间上恒定频率的时间信号,该时间扭曲操作可包括时间扭曲重新采样。

[0131] 图 3i 示出了未扭曲时间信号 $a(t)$ 的自相关函数的图形表示。横坐标 360 描述了自相关延迟 τ ,纵坐标 361 描述了该自相关函数的量值。标记 362 描述了自相关函数 $R_{uw}(\tau)$ 的演进,作为自相关延迟 τ 的函数。如图 3i 所示,未扭曲时间信号 $a(t)$ 的自相关函数 R_{uw} 包括 $\tau = 0$ 的峰值(由信号 $a(t)$ 的能量反映),且当 $\tau \neq 0$ 时为很小值。

[0132] 图 3j 示出了时间扭曲时间信号 $a(t_w)$ 的自相关函数 R_{tw} 的图形表示。如图 3j 所示,自相关函数 R_{tw} 包括 $\tau = 0$ 时的峰值,而且还包括自相关延迟 τ 的其它值 τ_1 、 τ_2 、 τ_3 时的峰值。这些 τ_1 、 τ_2 、 τ_3 的附加峰值由时间扭曲的效果获得,以增加时间扭曲时间信号 $a(t_w)$ 的周期性。当与自相关函数 $R_{uw}(\tau)$ 相比时,该周期性由自相关函数 $R_{tw}(\tau)$ 的附加峰值反映。因此,当相比于原始音频信号的自相关函数时,时间扭曲音频信号的自相关函数的附加波峰(或波峰的增加的强度)的存在性可用作时间扭曲的效能(就比特率减少而言)的指示。

[0133] 图 3k 示出了能量压缩信息提供器 370 的示意框图,其被配置为接收该音频信号的时间扭曲时域表示,例如时间扭曲信号 234e、234k(其中忽略频谱域变换 234d、234j 及可选

择的分析加窗器 234b 及 234h), 且基于它们提供能量压缩信息 374, 该信息 374 可发挥能量压缩信息 372 的作用。图 3k 的能量压缩信息提供器 370 包括自相关计算器 371, 被配置为计算时间扭曲信号 $a(t_w)$ 在离散值 τ 的预定范围上的自相关函数 $R_{t_w}(\tau)$ 。能量压缩信息提供器 370 还包括自相关加法器 372, 被配置为将自相关函数 $R_{t_w}(\tau)$ 的多个值 (例如, 在离散值 τ 的预定范围上) 相加, 且提供所获得的和作为能量压缩信息 122、234m、234n。

[0134] 因此, 能量压缩信息提供器 370 允许提供指示时间扭曲效果的可靠信息, 而不需实际执行对输入音频信号 210 的时间扭曲时域版本的频谱域变换。因此, 有可能仅当基于由能量压缩信息提供器 370 提供的能量压缩信息 122、234m、234n, 发现时间扭曲实际产生增强的编码效率时, 才执行对输入音频信号 310 的时间扭曲版本的频谱域变换。

[0135] 综上所述, 根据本发明的实施例创建用于最终质量检测的概念。对作为结果的音调轮廓 (用于时间扭曲音频信号编码器中) 在其编码增益方面加以评估, 并且接受它或拒绝它。可以考虑若干关于频谱的稀疏度或编码增益的度量, 例如, 频谱平坦度量、逐频带分段频谱平坦度量、和 / 或感知熵。

[0136] 已经讨论了不同频谱压缩信息的使用, 例如, 频谱平坦度度的使用, 感知熵度度的使用, 及时域自相关度度的使用。然而, 仍存在示出了时间扭曲频谱中的能量压缩的其它度量。

[0137] 可以使用所有这些度量。优选地, 对于所有这些度量, 定义未扭曲与时间扭曲频谱的度量之间的比率, 且在编码器中针对该比率设置阈值, 以确定已获得的时间扭曲轮廓在编码中是否有利。

[0138] 可将所有这些度量应用于全帧中, 在该帧中该音调轮廓仅三分之一是新的 (其中, 例如, 该音调轮廓的三个部分与该全帧相关联), 或优选地仅对于部分信号应用所有这些度量, 对于部分信号, 使用例如以位于 (各自) 信号部分中心的低重叠窗口的变换来获得该新部分。

[0139] 自然, 一单一度量或上述度量的一合并可被使用, 如所希望的。

[0140] 图 4a 示出了一种用于基于音频信号提供时间扭曲激活信号的方法的流程图。图 4a 的方法 400 包括提供能量压缩信息的步骤 410, 该能量压缩信息描述该音频信号的时间扭曲变换频谱表示中的能量压缩。方法 400 还包括将该能量压缩信息与参考值相比较的步骤 420。方法 400 还包括取决于该比较的结果提供时间扭曲激活信号的步骤 430。

[0141] 方法 400 可由本文与提供时间扭曲激活信号相关描述的任何特征及功能来补充。

[0142] 图 4b 示出了一种用于对输入音频信号编码以获得该输入音频信号的编码表示的方法的流程图。方法 450 可选地包括基于输入音频信号提供时间扭曲变换频谱表示的步骤 460。方法 450 还包括提供时间扭曲激活信号的步骤 470。步骤 470 可以包括例如方法 400 的功能。因此, 可以提供该能量压缩信息, 使得该能量压缩信息描述输入音频信号的时间扭曲变换频谱表示中的能量压缩。方法 450 还包括步骤 480, 取决于时间扭曲激活信号, 使用新发现时间扭曲轮廓信息提供对输入音频信号的时间扭曲变换频谱表示的描述, 或使用标准 (不变) 时间扭曲轮廓信息提供对输入音频信号的非时间扭曲变换频谱表示的描述, 以包括在输入信号的编码表示中。

[0143] 方法 450 可由与输入音频信号的编码相关本文讨论的任何特征及功能来补充。

[0144] 图 5 示出了根据本发明的音频编码器的优选实施例, 其中, 实施本发明的若干方

面。将音频信号提供于编码器输入 500 处。该音频信号将一般是离散音频信号,该离散音频信号使用被称作正常采样率的采样率从模拟音频信号中导出。该正常采样率不同于在时间扭曲操作中产生的本地采样率,且输入 500 处的音频信号的正常采样率是导致由恒定时间部分分开的音频采样的恒定采样率。将该信号输入分析加窗器 502,在该实施例中,将分析加窗器 502 连接至窗口函数控制器 504。分析加窗器 502 连接至时间扭曲器 506。然而,取决于实施,可以在信号处理方向上将时间扭曲器 506 置于分析加窗器 502 之前。当要求时间扭曲特性用于块 502 的分析窗口化时,且当要在时间扭曲采样上而非未扭曲采样上执行该时间扭曲操作时,该实施是优选的。特别在如国际专利申请 PCT/EP2009/002118, Bernd Edler 等人的“Time Warped MDCT”所描述的基于 MDCT 的时间扭曲的上下文中。对于其它时间扭曲应用,如 L. Villemoes 在 2005 年 11 月提出的国际专利申请 PCT/EP2006/010246, “Time Warped Transform Coding of Audio Signals”中描述的,时间扭曲器 506 与分析加窗器 502 之间的布置可按照需求来设置。此外,提供时间 / 频率转换器 508 用于执行时间扭曲音频信号到频谱表示的时间 / 频率转换。可将该频谱表示输入至 TNS(时域噪声修整)级 510,其作为输出 510a 提供 TNS 信息,且作为输出 510b 提供频谱残余值。将输出 510b 耦合至量化器及编码器块 512,该量化器及编码器块 512 可由感知模型 514 来控制,用于量化信号,使得将该量化噪声隐藏在音频信号的感知屏蔽阈值之下。

[0145] 此外,图 5a 所示编码器包括时间扭曲分析器 516,可将其实施为音调追踪器,其在输出 518 处提供时间扭曲信息。线 518 上的信号可以包括时间扭曲特性、音调特性、音调轮廓,或由时间扭曲分析器分析的信号是谐波信号还是非谐波信号的信息。该时间扭曲分析器还可实施区别有发音语音与无发音语音的功能。然而,取决于实施,及是否实施了信号分类器 520,有发音 / 无发音判定也可由信号分类器 520 来完成。在此情况中,该时间扭曲分析器没必要必须执行相同的功能。将时间扭曲分析器输出 518 连接至包括窗口函数控制器 504、时间扭曲器 506、TNS 级 510、量化器与编码器 512 及输出接口 522 在内的功能组中的至少一个且优选地多于一个的功能。

[0146] 类似地,可以将信号分类器 520 的输出 522 连接至包括窗口函数控制器 504、TNS 级 510、噪声填充分析器 524 或输出接口 522 在内的功能组中的至少一个且优选地多于一个的功能。此外,还可以将时间扭曲分析器输出 518 连接至噪声填充分析器 524。

[0147] 虽然图 5a 示出了将分析加窗器输入 500 上的音频信号输入至时间扭曲分析器 516 及信号分类器 520 的情况,针对这些功能的输入信号也可取自分析加窗器 502 的输出,以及信号分类器的输入甚至可取自时间扭曲器 506 的输出、时间 / 频率转换器 508 的输出或 TNS 级 510 的输出。

[0148] 除在 526 处指示的由量化器编码器 512 输出的信号外,输出接口 522 接收 TNS 侧信息 510a、感知模型侧信息 528,其可包括编码形式的扩缩因子,针对更高级的时间扭曲侧信息的时间扭曲指示数据,诸如线 518 上的音调轮廓及线 522 上的信号分类信息。此外,噪声填充分析器 524 还可以在输出 530 上将噪声填充数据输出至输出接口 522 中。输出接口 522 被配置为在线 532 上产生编码音频输出数据,以发送至解码器,或存储在存储设备(如存储器设备)中。取决于实施,输出数据 532 可包括到输出接口 522 的所有输入,或如果该信息不被对应的具有减少功能的解码器所需要,或如果该信息由于经由不同发送信道的发送已在该解码器处可用时,可包括更少信息。

[0149] 除了图 5a 中创造性编码器中所示的附加功能外,可以如 MPEG-4 标准中所详细定义的来实施图 5a 所示编码器,这些附加功能由相对于 MPEG-4 标准具有高级功能的窗口函数控制器 504、噪声填充分析器 524、量化器编码器 512 及 TNS 级 510 来表示。在 AAC 标准(国际标准 13818-7)或 3GPP TS 26.403V7.0.0:Third generation partnership project; technical specification group services and system aspect;general audiocodec audio processing functions;enhanced AAC plus general audiocodec 中对其进行进一步描述。

[0150] 随后,讨论图 5b,其示出了用于对经由输入 540 接收的编码音频信号进行解码的音频解码器的优选实施例。该输入接口 540 作用以处理编码音频信号,使得从在线 540 上信号中提取信息的不同信息项。该信息包括信号分类信息 541、时间扭曲信息 542、噪声填充数据 543、扩缩因子 544、TNS 数据 545 及编码频谱信息 546。将该编码频谱信息输入至熵解码器 547,假如将图 5a 的块 512 中的编码器功能实施为相对应的编码器,如霍夫曼编码器或算术编码器,则熵解码器 547 可包括霍夫曼解码器或算术解码器。将该解码频谱信息输入至重新量化器 550 中,将该重新量化器 550 连接至噪声填充器 552。将噪声填充器 552 的输出输入至反 TNS 级 554 中,反 TNS 级 554 附加地接收线 545 上的 TNS 数据。取决于实施,可以用不同的顺序来应用噪声填充器 552 及 TNS 级 554,使得噪声填充器 552 操作于 TNS 级 554 输出数据上而不是在 TNS 输入数据上。此外,提供频率/时间转换器 556,其向时间解扭器 558 馈送。在信号处理链的输出处,如 560 所指示地应用合成加窗器,其优选地执行重叠/添加处理。时间解扭器 558 与合成级 560 的顺序可改变,但是,在优选实施例中,优选地执行如在 AAC 标准(AAC = 高级音频编码)中定义的基于 MDCT 的编码/解码算法。接着,由于重叠/添加步骤而产生的从一个块到下一个块的固有交叉衰落操作有利地用作处理链中最后的操作,使得有效地避免所有的块伪像(artifact)。

[0151] 此外,提供噪声填充分析器 562,其被配置为控制噪声填充器 552,且接收作为输入的时间扭曲信息 542 和/或信号分类信息 541,以及与重新量化频谱相关的信息(看情况)。

[0152] 优选地,将此后所述的全部功能一起应用于增强的音频编码器/解码器方案中。然而,还可以彼此独立地应用此后所述功能,即,使得在特定编码器/解码器方案中实施仅一个或一组但非全部这些功能。

[0153] 随后,详细描述本发明的噪声填充方面。

[0154] 在实施例中,由图 5a 中时间扭曲/音调轮廓工具 516 提供的附加信息有利地用于控制其它编解码工具,且具体地,用于控制由编码器侧噪声填充分析器 524 所实施的和/或由解码器侧噪声填充分析器 562 及噪声填充器 552 实施的噪声填充工具。

[0155] AAC 框架中的若干编码器工具(如噪声填充工具)由音调轮廓分析收集的信息和/或由信号分类器 520 提供的信号分类的附加知识来控制。

[0156] 发现的音调轮廓以清晰谐波结构来指示信号段,所以谐波线之间的噪声填充可能减少感知质量,特别是语音信号上的,因此当发现音调轮廓时,减少噪声级别。否则,在部分音调之间会有噪声,此与模糊频谱的增加量化噪声具有相同的效果。此外,可通过使用信号分类器信息来进一步对噪声级别减少量求精,所以,例如对于语音信号将不会有噪声填充,且将对具有强谐波结构的一般信号应用中度噪声填充。

[0157] 大体上,噪声填充器 552 有助于向解码频谱插入频率线,其中,已经从编码器向解码器发送了零,即图 5a 的量化器 512 已经将频谱线量化为零。当然,将频谱线量化为零大大减少了已发送信号的比特率,且理论上,当这些频谱线低于由感知模型 514 确定的感知屏蔽阈值之下时,这些(小)频谱线的消除是不可听见的。然而,已发现可包括许多相邻频谱线的这些“频谱孔”导致相当不自然的声音。因此,提供噪声填充工具以在线由编码器侧量化器量化为零的位置处插入频谱线。这些频谱线可具有随机振幅或相位,且使用如图 5a 所示在编码器侧确定的噪声填充度量,或取决于图 5b 所示在解码器侧由可选块 562 确定的度量来扩缩这些解码器侧合成频谱线。因此,图 5a 中的噪声填充分析器 524 被配置为用于对于该音频信号的时帧,估计被量化为零的音频值的能量的噪声填充度量。

[0158] 在本发明的实施例中,用于对线 500 上的音频信号编码的音频编码器包括量化器 512,被配置为量化音频值,此外量化器 512 被配置为将在量化阈值之下的音频值量化为零。该量化阈值可以是基于阶的量化器的第一阶,用于判定是否将特定音频值量化为零(即,量化索引零),还是被量化为一(即,指示音频值在该第一阈值之上的量化索引一)。虽然将图 5a 的量化器示为执行频域值的量化,在备选实施例中该量化器还可用于量化时域值,其中,在时域而非在频域中执行噪声填充。

[0159] 将噪声填充分析器 524 实施为噪声填充计算器,用于估计该音频信号的时帧的由量化器 512 量化为零的音频值的能量的噪声填充度量。此外,音频编码器包括图 6a 所示的音频信号分析器 600,被配置为用于分析音频信号的时帧具有谐波特性还是语音特性。信号分析器 600 可包括例如图 5a 的块 516 或图 5a 的方块 520,或可包括用于分析信号是谐波信号还是语音信号的任何其它设备。由于将时间扭曲分析器 516 实施为总是寻找音调轮廓,且因为音调轮廓的存在性指示该信号的谐波结构,可将图 6a 中的信号分析器 600 实施为音调追踪器或时间扭曲分析器的时间扭曲轮廓计算器。

[0160] 该音频编码器附加地包括图 6a 所示的噪声填充级别操纵器 602,其输出经操纵的噪声填充度量/级别,要向图 5a 的 530 处所指示的输出接口 522 输出该经操纵的噪声填充度量/级别。噪声填充度量操纵器 602 被配置为取决于音频信号的谐波或语音特性来操纵该噪声填充度量。音频编码器附加地包括输出接口 522,用于产生编码信号供发送或存储之用,该编码信号包括由线 530 上的块 602 输出的经操纵的噪声填充度量。该值对应于由图 5b 所示的解码器侧实施中的块 562 输出的值。

[0161] 如图 5a 及图 5b 所示,可在编码器中实施或在解码器中实施、或在这两个装置中实施噪声填充级别操纵。在解码器侧实施中,用于对编码音频信号解码的解码器包括输入接口 539,用于处理线 540 上的编码信号,以获得噪声填充度量,即线 543 上的噪声填充数据,及线 546 上的编码音频数据。该解码器附加地包括解码器 547 及重新量化器 550 用于产生重新量化的数据。

[0162] 此外,解码器包括信号分析器 600(图 6a),可在图 5b 的噪声填充分析器 562 中实施为用于检索该音频数据的时帧具有谐波还是语音特性的信息。

[0163] 此外,提供噪声填充器 552 以产生噪声填充音频数据,其中噪声填充器 552 被配置为响应于以下各项来产生噪声填充数据:经由编码信号发送且由线 543 上的输入接口产生的噪声填充度量,以及由信号分析器 516 和/或 550 在编码器侧定义的或项 562 在解码器侧定义的,经由处理及解释指示特定时帧是否受到时间扭曲处理的时间扭曲信息 542 的音

频数据的谐波或语音特性。

[0164] 此外,该解码器包括处理器,用于处理重新量化的数据及噪声填充音频数据,以获得解码音频信号。该处理器可看情况包括图 5b 中的项 554、556、558、560。此外,取决于编码器/解码器算法的特定实施,该处理器可包括例如在时域编码器(如 AMR WB+ 编码器或其它语音编码器)中提供的其它处理块。

[0165] 因此,仅通过计算简单噪声度量,及通过基于谐波/语音信息来操纵该噪声度量,及通过发送可由解码器以简单方式应用的已正确经操纵的噪声填充度量,可在该编码器侧实施该创造性噪声填充操纵。备选地,可从编码器向解码器发送该未经操纵的噪声填充度量,且该解码器将进而分析是否已经对音频信号的实际时帧进行了时间扭曲,即,具有谐波还是语音特性,使得该噪声填充度量的实际操纵发生在解码器侧。

[0166] 随后,讨论图 6b 以解释用于操纵噪声级别估计的优选实施例。

[0167] 在第一实施例中,当该信号不具有谐波或语音特性时,应用正常噪声级别。这是当没有应用时间扭曲的情况。此外,当提供信号分类器时,则区分语音与无语音的信号分类器将针对该情况指示无语音,其中,时间扭曲无效,即,没有发现音调轮廓。

[0168] 然而,当时间扭曲有效时,即,当发现指示谐波内容的音调轮廓时,则将该噪声填充级别操纵为低于正常情况。当提供附加信号分类器时且该信号分类器指示语音时,同时当时间扭曲信息指示音调轮廓时,则以信号方式发送较低或甚至为零的噪声填充级别。因此,图 6a 的噪声填充级别操纵器 602 将经操纵的噪声级别减少至零,或至少为比图 6b 中指示的低值更低的值。优选地,该信号分类器附加地具有如图 6b 左边指示的有发音/无发音检测器。在有发音语音的情况中,以信号方式发送或应用很低的或零噪声填充级别。然而,在无发音语音的情况中,由于没有发现音调,时间扭曲指示不指示时间扭曲处理,但是信号分类器以信号方式发送语音内容,则不操纵该噪声填充度量,但是应用正常噪声填充级别。

[0169] 优选地,该音频信号分析器包括音调追踪器用于产生该音调的指示,如音频信号的时帧的音调轮廓或绝对音调。然后,该操纵器被配置为用于当发现音调时,减少该噪声填充度量,且当未发现音调时,不减少该噪声填充度量。

[0170] 如图 6a 所示,当应用于解码器侧时,信号分析器 600 不像音调追踪器或有发音/无发音检测器那样执行实际信号分析,但是该信号分析器解析编码音频信号,以提取时间扭曲信息或信号分类信息。因此,可在图 5b 解码器的输入接口 539 中实施信号分析器 600。

[0171] 随后将参考图 7a-7e 来讨论本发明的另一实施例。

[0172] 对于有发音语音部分在相对安静信号部分后开始的语音的起始点而言,块切换算法可将其分类成攻击(attack),且可以针对该特定帧选择短块,同时在具有清晰谐波结构的信号段上损失编码增益。因此,该音调追踪器的有发音/无发音分类用于检测有发音起始,且避免该块切换算法指示围绕发现的起始点的瞬变攻击。该特征也可与信号分类器耦合以防止语音信号上的块切换,且允许它们针对所有的其它信号。此外,该块切换的更精细控制可通过不仅允许或不允许攻击检测,还使用基于有发音起始及信号分类信息的针对攻击检测的可变阈值。此外,该信息可用于检测类似上述有发音起始的攻击,但是不切换至短块,而是使用具有短重叠的长窗口,具有短重叠的长窗口保留了优选频谱解析度,但是减少了前回声和后回声可能出现的时间区域。图 7d 示出了未调整的典型行为,图 7e 示出了调整的两种不同可能性(防止及低重叠窗口)。

[0173] 根据本发明的实施例的音频编码器操作以产生音频信号,如由图 5a 的输出接口 522 输出的信号。该音频编码器包括音频信号分析器,如图 5a 的时间扭曲分析器 516 或信号分类器 520。大体上,该音频信号分析器分析该音频信号的时帧具有谐波还是语音特性。为此,图 5a 的信号分类器 520 可包括有发音 / 无发音检测器 520a 或语音 / 无语音检测器 520b。虽然图 7a 未示出,可提供取代项 520a 及 520b,或与这些功能一起提供的可包括音调追踪器在内的时间扭曲分析器,如图 5a 的时间扭曲分析器 516。此外,该音频编码器包括窗口函数控制器 504,用于取决于由音频信号分析器确定的音频信号的谐波或语音特性,来选择窗口函数。加窗器 502 进而窗口化该音频信号,或取决于特定实施,使用已选择窗口函数窗口化时间扭曲音频信号,以获得窗口型帧。该窗口帧接着还由处理器处理,以获得编码音频信号。该处理器可包括图 5a 所示的项 508、510、512,或众所周知的音频编码器(如基于变换的音频编码器),或包括 LPC 滤波器的基于时域的音频编码器(如语音编码器及,具体地根据 AMR-WB+ 标准所实施的语音编码器)的或多或少的功能。

[0174] 在优选实施例中,窗口函数控制器 504 包括瞬变检测器 700,用于检测该音频信号中的瞬变,其中该窗口函数控制器被配置为当检测到瞬变,且音频信号分析器未发现谐波或语音特性时,将针对长块的窗口函数切换至针对短块的窗口函数。然而,当检测到瞬变,且音频信号分析器发现谐波或语音特性时,则窗口函数控制器 504 不切换至针对短块的窗口函数。如图 7a 的 701 及 702 示出了窗口函数输出,其指示当未获得没有瞬变的长窗口及瞬变检测器检测到瞬变时的短窗口。图 7d 示出了由众所周知的 AAC 编码器执行的该正常步骤。在有发音起始的位置上,瞬变检测器 700 检测到能量从一个帧到下一帧的增加,且因此,从长窗口 710 切换至短窗口 712。为了适应该切换,使用长终止窗口 714,其具有第一重叠部分 714a、非混叠部分 714b、第二较短重叠部分 714c、及在点 716 和由 2048 个采样指示的时间轴上点之间扩展的零值部分。然后,执行在 712 处指示的短窗口的序列,接着由具有与未示于图 7d 中的下一个长窗口重叠的长重叠部分 718a 的长起始窗口 718 来结束。此外,该窗口具有非混叠部分 718b、短重叠部分 718c 及在点 720 和时间轴上在直到第 2048 点之间扩展的零值部分。该部分是零值部分。

[0175] 通常,为了避免会在该瞬变事件前的帧中发生前回声,至短窗口的切换是有用的,该帧是有发音起始的位置,或一般而言,是该语音的开始或具有谐波内容的信号的开始的位置。大体上,当音调追踪器确定信号具有音调时,该信号具有谐波内容。备选地,存在其它谐波度量,如音调度量,其在特定最小级别之上并具有突出波峰彼此处于谐波关系的特性。存在多个其他技术用于确定信号是否为谐波的。

[0176] 短窗口的缺点是减少了频率解析度,因为增加了时间解析度。对于语音的高质量编码,且具体地,对于有发音语音部分或具有强谐波内容的部分的高质量编码,需要好的频率解析度。因此,在 516、520 或 520a、520b 处所示的音频信号分析器操作以向瞬变检测器 700 输出停用信号,使得当检测到有发音语音段或具有强谐波特性的信号段时,阻止切换至短窗口。这确保了对于编码这种信号部分,维持了高频解析度。这是一方面前回声与另一方面对于语音信号的音调或谐波无语音信号的音调的高质量且高解析度编码之间的折中。已发现当与将发生的任何前回声相比较时,没有对谐波频谱进行精确编码更加令人烦恼。为了进一步减少前回声,TNS 处理有利于这种情况,将通过图 8a 与 8b 来讨论该 TNS 处理。

[0177] 在图 7b 所示的备选实施例中,音频信号分析器包括有发音 / 无发音和 / 或语音 /

无语音检测器 520a、520b。然而，窗口函数控制器中包括的瞬变检测器 700 未如图 7a 所示完全激活 / 停用，但是使用阈值控制信号 704 来控制瞬变检测器中包括的阈值。在该实施例中，瞬变检测器 700 被配置为用于确定该音频信号的定量特性，且用于将该定量特性与可控阈值相比较，其中当该定量特性具有与可控阈值的预定关系时，检测到瞬变。该定量特性可以是指示从一个块到下一个块的能量增加的数量，且该阈值可以是特定阈值能量增加。当从一个块到下一个的能量增加高于阈值能量增加时，那么检测到瞬变，使得，在该情况中，预定关系是“高于”关系。在其它实施例中，该预定关系也可以是“低于”关系，例如当该定量特性是反向能量增加时。在图 7b 的实施例中，控制该可控阈值，使得当该音频信号分析器已发现谐波或语音特性时，减少切换至针对短块的窗口函数的可能性。在能量增加实施例中，阈值控制信号 704 将导致阈值的增加，使得仅当从一个块到下一个的能量增加是特别高的能量增加时，至短块的切换才发生。

[0178] 在备选实施例中，来自有发音 / 无发音检测器 520a 或语音 / 无语音检测器 520b 的输出信号还可用于用如下方法控制窗口函数控制器 504：执行切换至比针对短块的窗口函数更长的窗口函数，而不是在语音起始处切换至短块。该窗口函数确保比短窗口函数更高的频率解析度，但是具有比长窗口函数更短的长度，使得获得在一方面的前回声与另一方面的充分的频率解析度之间的良好折中。在备选实施例中，可以如图 7e 中 706 处的虚线所示，执行到具有较小重叠的窗口函数的切换。窗口函数 706 具有如长块的 2048 个采样的长度，但是该窗口具有零值部分 708 及非混叠部分 710，使得获得从窗口 706 到相对应窗口 707 的短重叠长度 712。窗口函数 707 同样具有与窗口函数 710 类似的区域 712 的左边的零值部分，及区域 712 右边的非混叠部分。该低重叠实施例，有效地导致较短时间长度，用于减少由于窗口 706 与 707 的零值部分而产生的前回声，但是另一方面具有由于重叠部分 714 及非混叠部分 710 而产生的充分长度，使得维持了充足的频率解析度。

[0179] 在由 AAC 编码器实施的优选 MDCT 实施中，维持特定重叠提供了如下附加优势：在解码器侧，可执行重叠 / 添加处理，其意味着执行块之间的交叉衰落。这有效地避免了区块伪像。此外，该重叠 / 添加特征 提供该交叉衰落特性，而不增加比特率，即，获得关键采样的交叉衰落。在规律的长窗口或短窗口中，该重叠部分是由重叠部分 714 指示的 50% 的重叠。在窗口函数为 2048 个采样长的实施例中，该重叠部分是 50%，即 1024 个采样。具有较短重叠的窗口函数优选地少于 50%，且在图 7e 实施例中，仅为 128 个采样，是整个窗口长度的 1/16，该较短重叠用于有效地窗口化语音起始或谐波信号的起始。优选地，使用在全部窗口函数长度的 1/4 与 1/32 之间的重叠部分。

[0180] 图 7c 示出了该实施例，其中示例性有发音 / 无发音检测器 520a 控制窗口函数控制器 504 中包括的窗口形状选择器，以选择在 749 处指示的具有短重叠的窗口形状，或选择在 750 处指示的具有长重叠的窗口形状。当有发音 / 无发音检测器 500a 在 751 处发出有发音检测信号时，实施对这两个形状之一的选择，其中，用于分析的音频信号可以是图 5a 的输入 500 处的音频信号，或是预处理音频信号（如时间扭曲信号或已受到任何其它预处理功能的音频信号）。优选地，当窗口函数控制器中包括的瞬变检测器将检测到瞬变，且如通过图 7a 所讨论的将命令从长窗口函数切换到短窗口函数时，图 5a 的窗口函数控制器 504 中包括的图 7c 中的窗口形状选择器 504 仅使用信号 751。

[0181] 优选地，将该窗口函数切换实施例与通过图 8a 和 8b 所讨论的时域噪声修整实施

例结合。然而,也可实施 TNS(时域噪声修整)实施例,而不需要块切换实施例。

[0182] 时间扭曲 MDCT 的频谱能量压缩性质还影响时域噪声修整 (TNS) 工具,因为对于时间扭曲帧,尤其是对于一些语音信号, TNS 增益趋于减少。然而需要激活 TNS,以例如在不需块切换,但是语音信号的时间包络显示出快速改变的情况下,减少有发音起始或偏移(参见块切换调整)的前回声。一般地,编码器使用某个度量以查看 TNS 的应用对特定帧是否有成效,例如当应用至频谱时 TNS 滤波器的预测增益。所以可变 TNS 增益阈值是优选的,其对具有有效音调轮廓的片段更低,因此确保 TNS 对这种类似有发音起始的关键信号部分更经常地有效。当用其它工具时,还可以通过考虑信号分类来加以补充。

[0183] 根据本实施例用于产生音频信号的音频编码器包括可控时间扭曲器,如用于对音频信号进行时间扭曲以获得时间扭曲音频信号的时间扭曲器 506。此外,提供了用于将至少一部分时间扭曲音频信号转换为频谱表示的时间/频率转换器 508。时间/频率转换器 508 优选地实施如来自众所周知的 AAC 编码器的 MDCT 变换,但是该时间/频率转换器还可执行任何其它种类的变换,如 DCT、DST、DFT、FFT 或 MDST 变换,或可包括滤波器组,如 QMF 滤波器组。

[0184] 此外,该编码器包括时域噪声修整级 510,用于根据时域噪声修整控制指令来执行对频谱表示的频率的预测滤波,其中当该时域噪声修整控制指令不存在时,不执行该预测滤波。

[0185] 此外,该编码器包括时域噪声修整控制器,用于基于频谱表示来产生时域噪声修整控制指令。

[0186] 具体地,该时域噪声修整控制器被配置为用于当频谱表示基于时间扭曲信号上时,增加对频率执行预测滤波的可能性,或用于当频谱表示不基于时间扭曲信号上时,减少对频率执行预测滤波的可能性。通过图 8 讨论了该时域噪声修整控制器的细节。

[0187] 该音频编码器附加地包括处理器,用于对频率的预测滤波的结果的进一步处理,以获得编码信号。在实施例中,该处理器包括图 5a 所示的量化器编码器级 512。

[0188] 在图 8 中详细说明了图 5a 所示的 TNS 级 510。优选地,级 510 中包括的时域噪声修整控制器包括 TNS 增益计算器 800、随后连接的 TNS 判定器 802 及阈值控制信号产生器 804。取决于来自时间扭曲分析器 516 或信号分类器 520 或两者之一的信号,该阈值控制信号产生器 804 向 TNS 确定器输出阈值控制信号 806。TNS 判定器 802 具有可控阈值,其根据阈值控制信号 806 而增加或减少。在本实施例中,在 TNS 判定器 802 中的阈值是 TNS 增益阈值。当由块 800 输出的实际计算的 TNS 增益超出阈值时,则 TNS 控制指令要求作为输出的 TNS 处理,而在其它情况中,当 TNS 增益低于 TNS 增益阈值时,不输出 TNS 指令,或输出指示该 TNS 处理没用且在该特定时帧中将不执行 TNS 处理的信号。

[0189] TNS 增益计算器 800 接收作为输入的从该时间扭曲信号导出的频谱表示。一般地,时间扭曲信号将具有较低的 TNS 增益,但是另一方面,由于时域中时域噪声修整特征而产生的 TNS 处理在该特定情况中是受益者,其中,存在已经受到时间扭曲操作的有发音/谐波信号。另一方面,TNS 处理在 TNS 增益很低的情况中没用,意指线 510b 上的 TNS 残余信号具有与 TNS 级 510 之前的信号相同的或更高的能量。在线 510d 上 TNS 残余信号的能量稍微低于 TNS 级 510 之前的能量的情况中,该 TNS 处理也可能不具优势,因为由于量化器/熵编码器级 512 有效使用的信号中稍小的能量而产生的比特减少小于由图 5a 中 510a 处指示

的 TNS 侧信息的必要发送所引入的比特增加。虽然一个实施例针对所有帧在 TNS 处理上自动切换,其中,时间扭曲信号是由来自块 516 的音调信息或来自块 520 的信号分类器信息所指示的输入,优选实施例同样维持停用 TNS 处理的可能性,但仅当该增益确实很低或至少低于没有处理谐波 / 语音信号的情况。

[0190] 图 8b 示出了由阈值控制信号产生器 804/TNS 判定器 802 实施三个不同阈值设置的实施。当音调轮廓不存在时,且当信号分类器指示无发音语音或没有语音时,则将 TNS 判定阈值设置在需要相对高的 TNS 增益用于激活 TNS 的正常状态中。然而,当检测到音调轮廓,但是信号分类器指示无语音或有发音 / 无发音检测器检测到无发音语音时,则将 TNS 判定阈值设置为较低级别,意指甚至当由图 8a 的块 800 计算相对低的 TNS 增益时,无论如何也激活 TNS 处理。

[0191] 在检测到有效音调轮廓且发现有发音语音的情况中,则将 TNS 判定阈值设置为相同的较低值,或设置为甚至更低的状态,使得即使很小 TNS 增益也足以激活 TNS 处理。

[0192] 在实施例中,TNS 增益控制器 800 被配置为当音频信号受到对频率的预测滤波时,估计在比特率或质量上的增益。TNS 判定器 802 将该估计增益与判定阈值进行比较,且当已估计增益与该确定阈值处于预定关系时,由块 802 来输出有利于预测滤波的 TNS 控制信息,其中预定关系可以是“高于”关系,例如对于反向 TNS 增益也可以是“低于”关系。正如所讨论的,时域噪声修整控制器还被配置为优选地使用阈值控制信号 806 来改变判定阈值,使得对于相同的已估计增益,当频谱表示基于时间扭曲音频信号时,激活预测滤波,当频谱表示不基于时间扭曲音频信号时,不激活预测滤波。

[0193] 通常,有发音语音将显示音调轮廓,且无发音语音诸如摩擦音或齿擦音不显示音调轮廓。然而确实存在无语音信号,尽管语音检测器没有检测到语音,但其具有强谐波内容,因此具有音调轮廓。此外,存在特定的基于音乐的语音或基于语音的音乐,由音频信号分析器(例如图 5a 的 516)确定其具有谐波内容,但是信号分类器 520 不将其检测为语音信号。在这种情况下,也可应用针对有发音语音信号的所有处理操作,且也将产生优势。

[0194] 随后,通过用于对音频信号编码的音频编码器来描述本发明的另一优选实施例。该音频编码器在带宽扩展的上下文中特别有用,且在独立编码器应用中也是有用的,在独立编码器应用中,音频编码器被设置为对特定数目的线编码,以获得特定带宽限制 / 低通滤波操作。在未时间扭曲应用中,通过选择特定预定数目线的带宽限制将导致恒定带宽,因为该音频信号的采样频率是恒定的。然而,在执行如图 5a 的块 506 的时间扭曲处理的情况下,依靠固定数目线的编码器将导致变化带宽,该变化的带宽引入不仅可由经过训练的收听者感知且可由未经训练收听者感知的很强的伪像。

[0195] AAC 核心编码器通常对固定数目的线编码,将所有其它在最大线之上的设为零。在该未扭曲情况中,这导致具有恒定截止频率的低通效应,且因此导致解码 AAC 信号的恒定带宽。在时间扭曲的情况下,带宽由于本地采样频率(与本地时间扭曲轮廓相关)的变化而变化,导致可听见的伪像。可通过取决于本地采样频率来适当地选择核心编码器中要编码的线的数目(与本地时间扭曲轮廓及其获得的平均采样率相关),使得在解码器中对所有帧的时间重新扭曲之后获得恒定平均带宽,来减少该伪像。附加好处是编码器中的比特节约。

[0196] 根据该实施例的音频编码器包括时间扭曲器 506,用于使用可变时间扭曲特性将

音频信号时间扭曲。此外,提供了用于将时间扭曲音频信号转换至具有一定数目频谱系数的频谱表示的时间/频率转换器 508。此外,使用用于处理可变数目的频谱系数以产生编码音频信号的处理器,其中,包括图 5a 的量化器/编码器块 512 的该处理器被配置为基于帧的时间扭曲特性,针对音频信号的帧设置一定数目的频谱系数,使得减少或消除帧与帧之间的已处理数目的频谱系数所表示的带宽变化。

[0197] 由块 512 实施的处理器包括控制器 1000,用于控制这些数目的线,控制器 1000 的结果是,相对于被编码而没有任何时间扭曲的时帧的情况所设置的一定数目的线,在频谱的上端添加或丢弃特定可变数目的线。取决于实施,控制器 1000 可接收特定帧 1001 中的音调轮廓信息,和/或在 1002 处指示的帧中的本地平均采样频率。

[0198] 在图 9(a)至 9(e)中,右边图片示出了在帧上的特定音调轮廓的特定带宽情况,在对应的左边图片上示出了时间扭曲的该帧上的音调轮廓,且在中间图片中示出了时间扭曲之后的该帧上的音调轮廓,其中获得实质上恒定的音调特性。在时间扭曲后音调特性尽可能的恒定是时间扭曲功能的目标。

[0199] 带宽 900 示出了,当采用由图 5a 的时间/频率转换器 508 所输出或由 TNS 级 510 所输出的特定数目的线时,且当未执行时间扭曲操作时,即当如虚线 507 所指示的停用时间扭曲器 506 时,所获得的带宽。然而,当获得非恒定时间扭曲轮廓时,且当将该时间扭曲轮廓带至引起采样率增加的较高音调时(图 9(a)、(c)),该频谱的带宽相对于正常、未时间扭曲的情况减少。这意指必须增加针对该帧要发送的线的数目,以平衡该带宽损失。

[0200] 备选地,将音调带至图 9(b)或图 9(d)所示的较低恒定音调中导致采样率的减少。该采样率减少导致该帧的频谱相对于线性尺度的带宽增加,且必须相对于正常未时间扭曲情况下的线的数目值,使用删除或丢弃特定数目的线来平衡该带宽增加。

[0201] 图 9(e)示出了特殊情况,其中将音调轮廓带至中间级别,使得帧中的平均采样频率与没有任何时间扭曲的采样频率相同,而不是执行时间扭曲操作。因此,尽管执行该时间扭曲操作,该信号的带宽不受影响,且可处理针对没有时间扭曲的正常情况所使用简单数目的线。从图 9,显而易见地,执行时间扭曲操作不一定影响带宽,但是对带宽的影响取决于音调轮廓及在帧中执行时间扭曲的方式。因此,优选地使用本地或平均采样率作为控制值。图 11 示出了该本地采样率的确定。图 11 的上部示出了具有等距采样值的时间部分。帧包括例如在较高图中由 T_n 指示的七个采样值。较低图示出了时间扭曲操作的结果,其中采样率增强发生。这意指该时间扭曲帧的时间长度小于未时间扭曲帧的时间长度。然而,因为要引入至时间/频率转换器的时间扭曲帧的时间长度是固定的,采样率增加的情况引起将时间信号的不属于由 T_n 所指示的帧的附加部分引入时间扭曲帧,如线 1100 所指示的。因此,时间扭曲帧覆盖有 T_{lin} 指示的音频信号的时间部分, T_{lin} 长于时间 T_n 。鉴于此,两条频率线之间的有效距离或线性域中的单一线的频率带宽(是该解析度的倒数值)减少,且当乘以减少的频率距离时,针对未时间扭曲情况设置的该数目的线 N_n 导致较小带宽,即,带宽减少。

[0202] 图 11 中未示出由时间扭曲器执行采样率减少的其它情况,在时间扭曲域中的帧的有效时间长度小于该未时间扭曲域中的时间长度,使得增加单一线的频率带宽或两个频率线之间的距离。现在对于正常情况,以线的数目 N_n 乘以增加的 Δf 将导致由于两个相邻频率系数之间的减少的频率解析度/增加的频率距离而增加的带宽。

[0203] 图 11 附加地示出了如何计算平均采样率 f_{SR} 。为此,确定两个时间扭曲采样之间的时间距离且采用倒数,该倒数被定义为两个时间扭曲采样之间的本地采样率。可在每对相邻采样之间计算这种值,且可计算算术平均值,且该值最终导致平均本地采样率,平均本地采样率优选地用于输入至图 10a 的控制器 1000 中。

[0204] 图 10b 示出了取决于本地采样频率来指示必须添加或丢弃多少线的图表,其中未扭曲情况的采样频率 f_N 与未时间扭曲情况的线的数目 N_N 定义了预期的带宽,对于一系列时间扭曲帧或一系列时间扭曲及未时间扭曲帧,应尽可能的将该带宽保持恒定。

[0205] 图 12b 示出了通过图 9、图 10b 及图 11 来讨论的不同参数之间的依赖性。基本上,当采样率(即平均采样率 f_{SR})相对于未时间扭曲情况减少时,必须删除线,而当采样率相对于正常采样率 f_N 增加时,必须添加线,以减少或优选地甚至尽可能地消除帧与帧之间的带宽变化。

[0206] 由这些数目的线 N_N 及采样率 f_N 产生的带宽优选地定义了音频编码器的交叉频率 1200,除源核心音频编码器之外,该音频编码器具有带宽扩展编码器(BWE 编码器)。如本领域中众所周知的,带宽扩展编码器仅以高比特率对频谱编码直到该交叉频率,且以低比特率对该高频带的频谱,即交叉频率 1200 与频率 f_{MAX} 之间的频谱进行编码,其中该低比特率一般甚至低于频率 0 与交叉频率 1200 之间的低频带所需的比特率的 1/10 或更少。此外,图 12a 示出了简单 AAC 音频编码器的带宽 BW_{AAC} ,其远高于该交叉频率。因此,不仅可丢弃线,也可加入线。此外,还示出了针对恒定数目线的带宽的变化取决于本地采样率 f_{SR} 。优选地,设置相对于正常情况的线的数目的要添加或要删除的线的数目,使得 AAC 编码数据的每一帧具有尽可能接近交叉频率 1200 的最大频率。因此,一方面避免了由于带宽减少,或由于低频带编码帧中在交叉频率之上的频率上发送信息的开销所产生的任何频谱孔。另一方面这增加了解码音频信号的质量,且另一方面减少了比特率。

[0207] 可在量化线之前(即在块 512 的输入处)执行,或可在量化之后执行,或取决于特定熵编码,也可在熵编码后执行相对于线的设置数目的线的实际添加,或相对于线的设置数目的线的删除。

[0208] 此外,优选地,将这些带宽变化带到最小级别,且甚至消除这些带宽变化,但是在其它实施中,与应用恒定数目的线而不管特定时间扭曲特性的情况相比较,通过取决于时间扭曲特性来确定线的数目以减少带宽变化提高了音频质量,且减少了所需要的比特率。

[0209] 尽管已在设备的上下文中描述一些方面,很明显,这些方面也表示对应方法的描述,其中块或设备对应于方法步骤或方法步骤的特征。类似地,在方法步骤的上下文中描述的方面也表示对应设备的对应块或项或特征的描述。

[0210] 取决于特定实施要求,可在硬件或软件中实施本发明的实施例。可使用数字存储介质,如磁盘、DVD、CD、ROM、PROM、EPROM、EEPROM 或 FLASH 存储器来执行该实施,该数字存储介质具有存储于其上的电子可读控制信号,该信号与(或能够与)可编程计算机系统配合,使得执行相应方法。根据本发明的一些实施例包括具有电子可读控制信号的数据载体,这些信号能够与可编程计算机系统配合,使得执行本文所述的方法之一。总体上,可以将本发明实施为具有程序代码的计算机程序产品,所述程序代码可操作于当该计算机程序产品在计算机上运行时,该程序代码执行这些方法之一。该程序代码可,例如存储于机器可读载体上。其它实施例包括存储于机器可读载体上的计算机程序,用于执行本文所描述的

方法之一。因此,换言之,该创造性方法的实施例是具有程序代码的计算机程序,当计算机程序运行于计算机上时,该程序代码用于执行本文所描述的方法之一。因此,该创造性方法的另一实施例是数据载体(或数字存储介质,或计算机可读介质),其包括记录于其上的计算机程序,用于执行本文所描述的这些方法之一。因此,该创造性方法的另一实施例是表示计算机程序的数据流或一系列信号,用于执行本文所描述的这些方法之一。该数据流或该系列信号可例如被配置为经由数据通信连接,例如经由互联网被传输。另一实施例包括处理装置,例如计算机,或可编程逻辑设备,被配置为或适于执行本文所描述的方法之一。另一实施例包括计算机,具有安装于其上的计算机程序,用于执行本文所描述的方法之一。在一些实施例中,可编程逻辑设备(例如现场可编程门阵列)可用于本文所描述的这些方法的一些或全部功能。在一些实施例中,现场可编程门阵列可与微处理器配合,以执行本文所描述的这些方法之一。

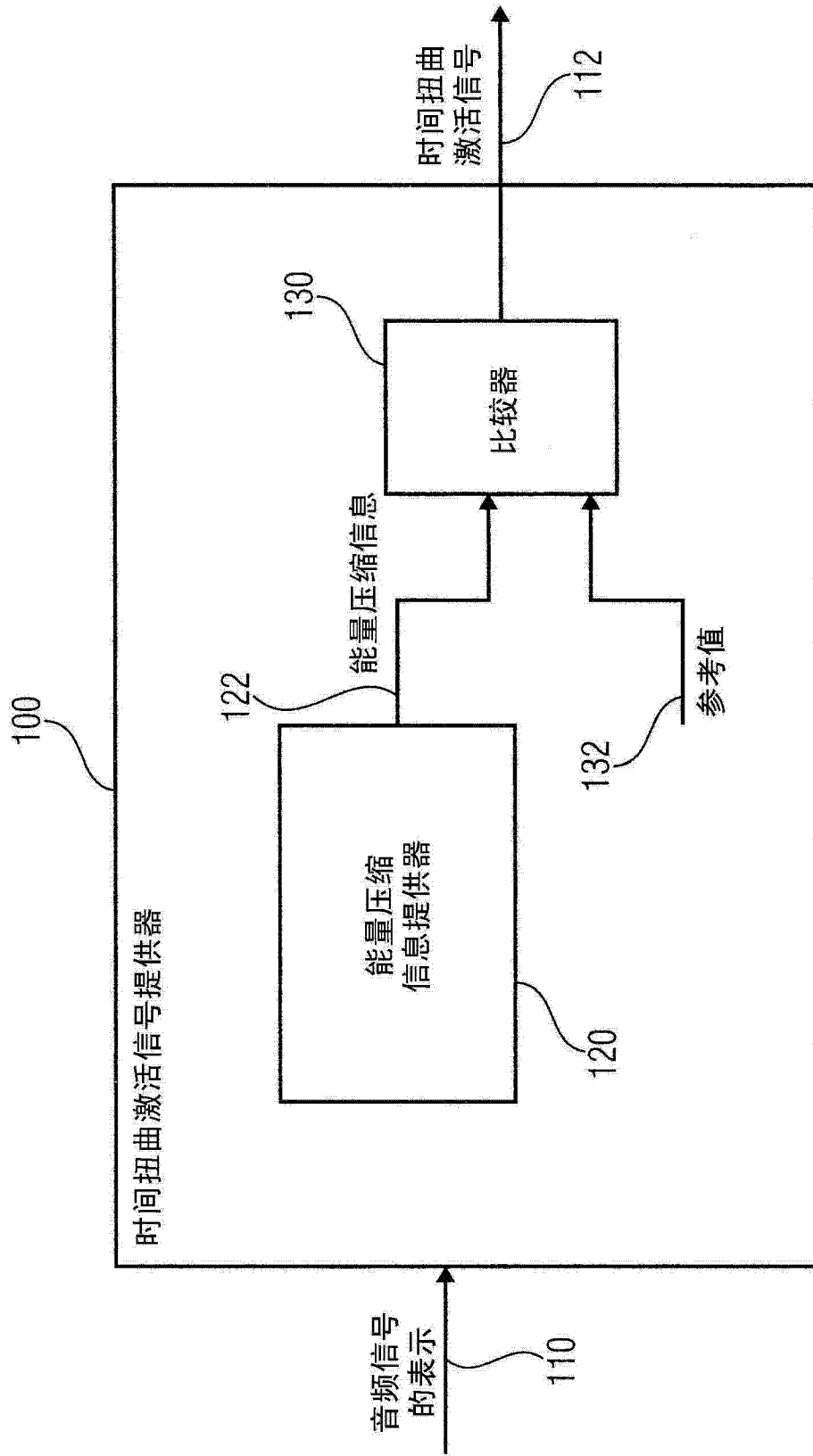


图 1

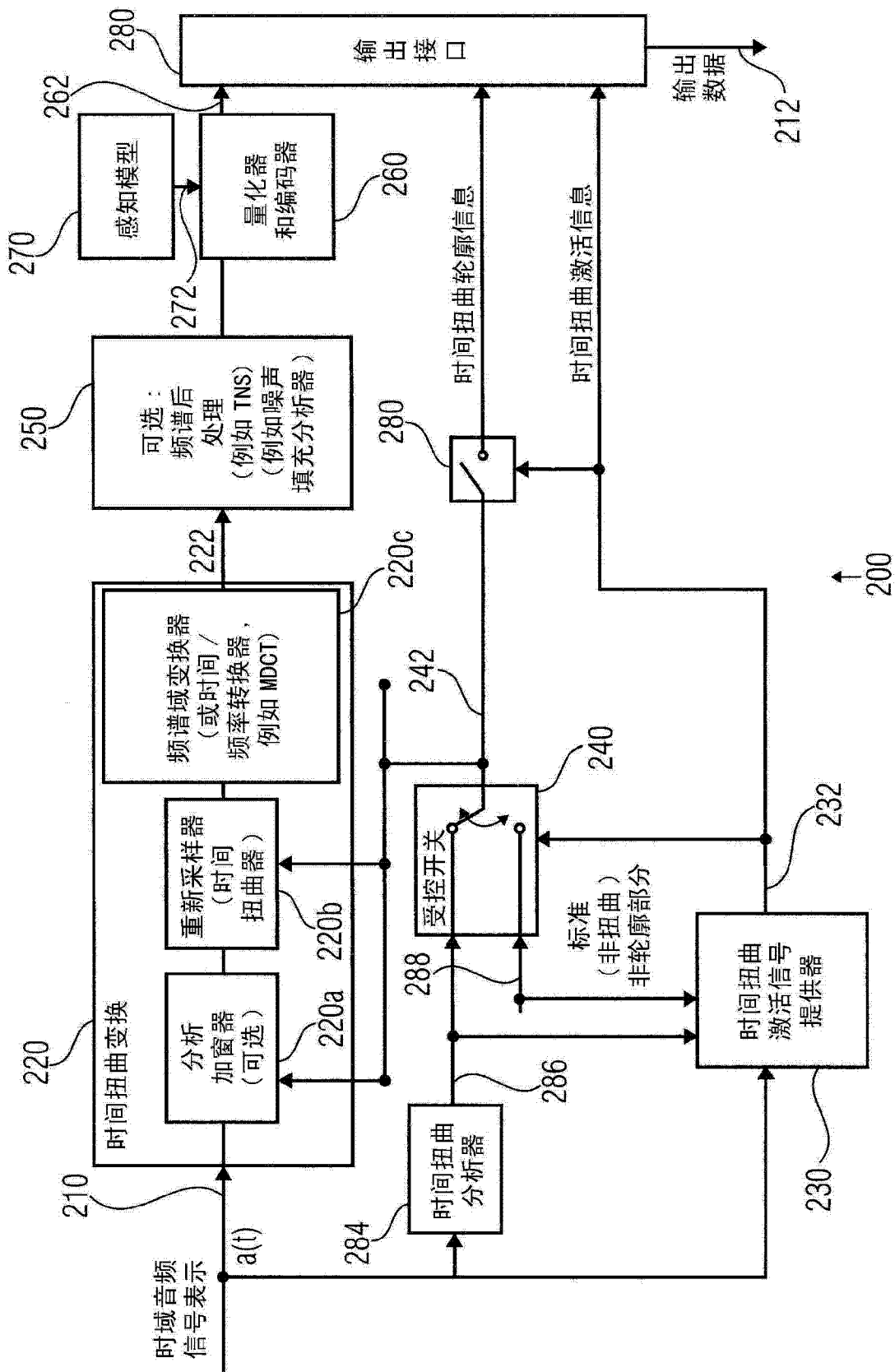


图 2A

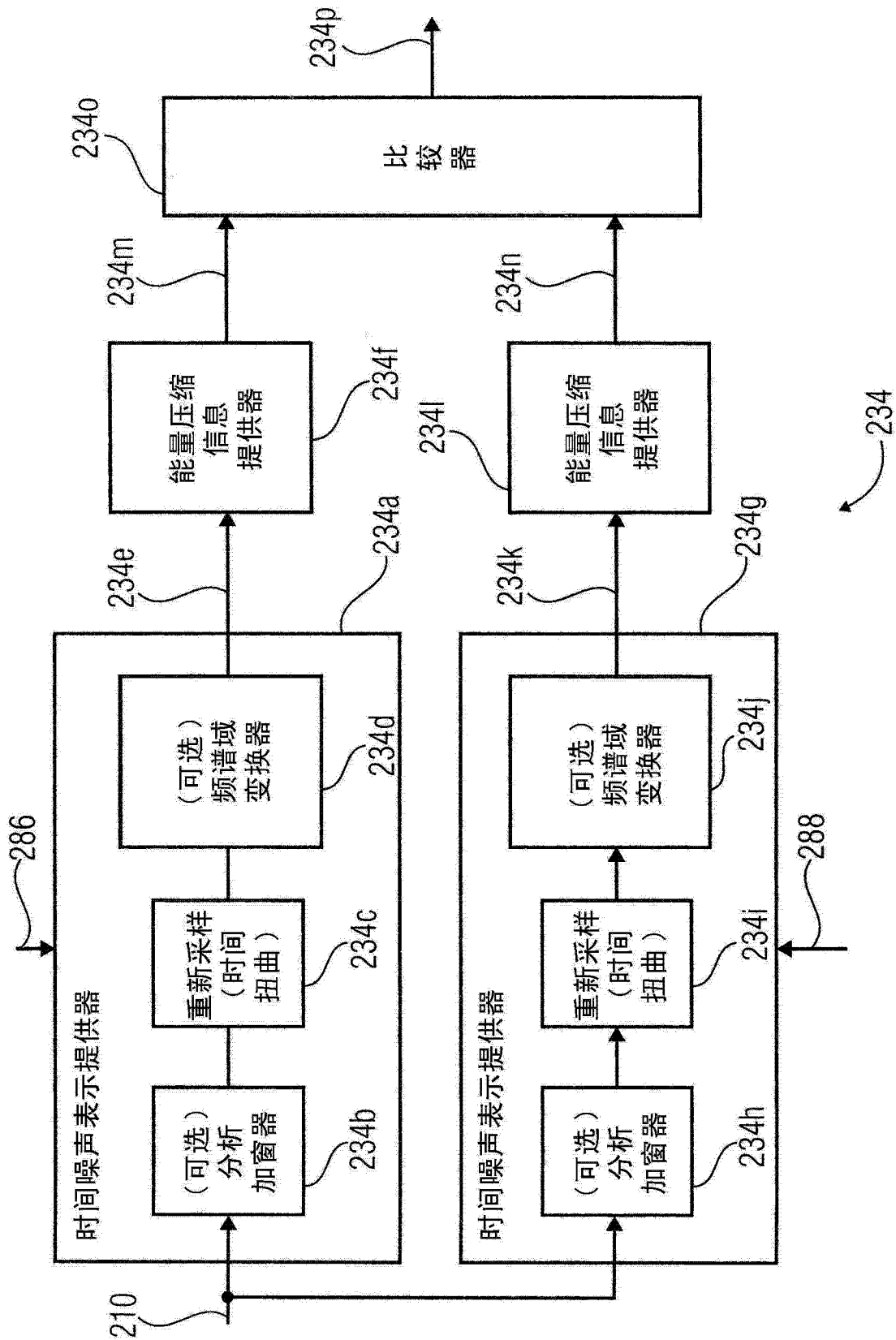


图 2B

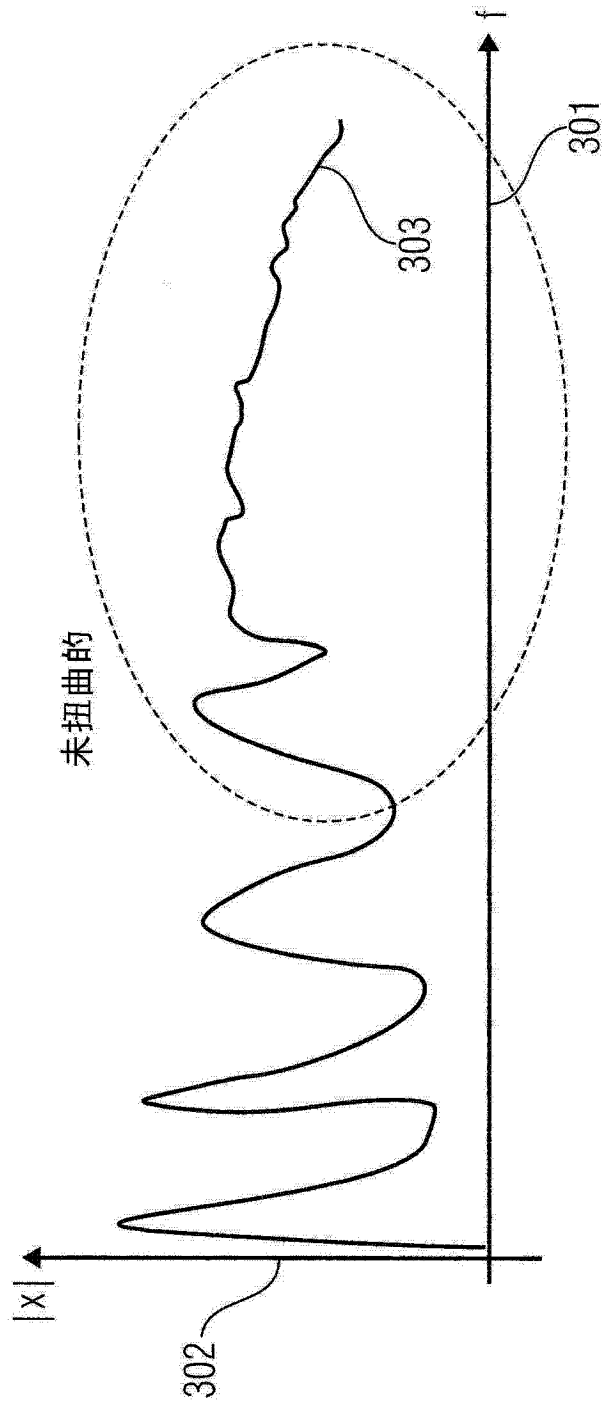


图 3A

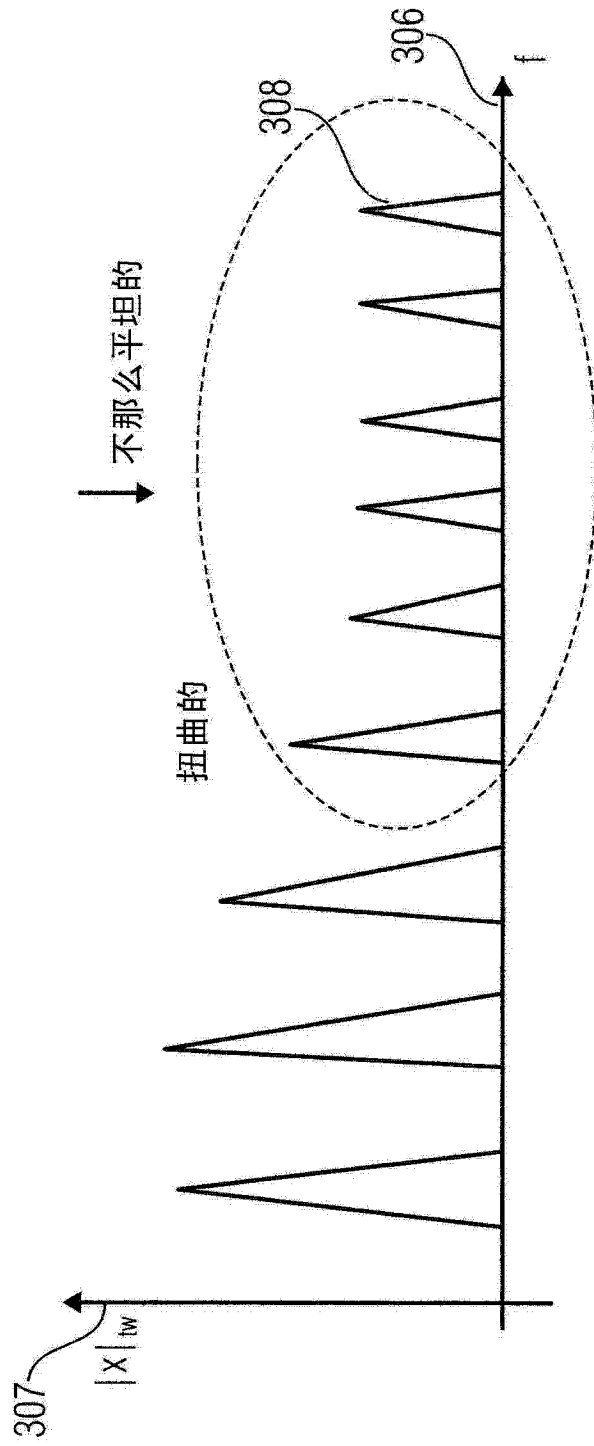


图 3B

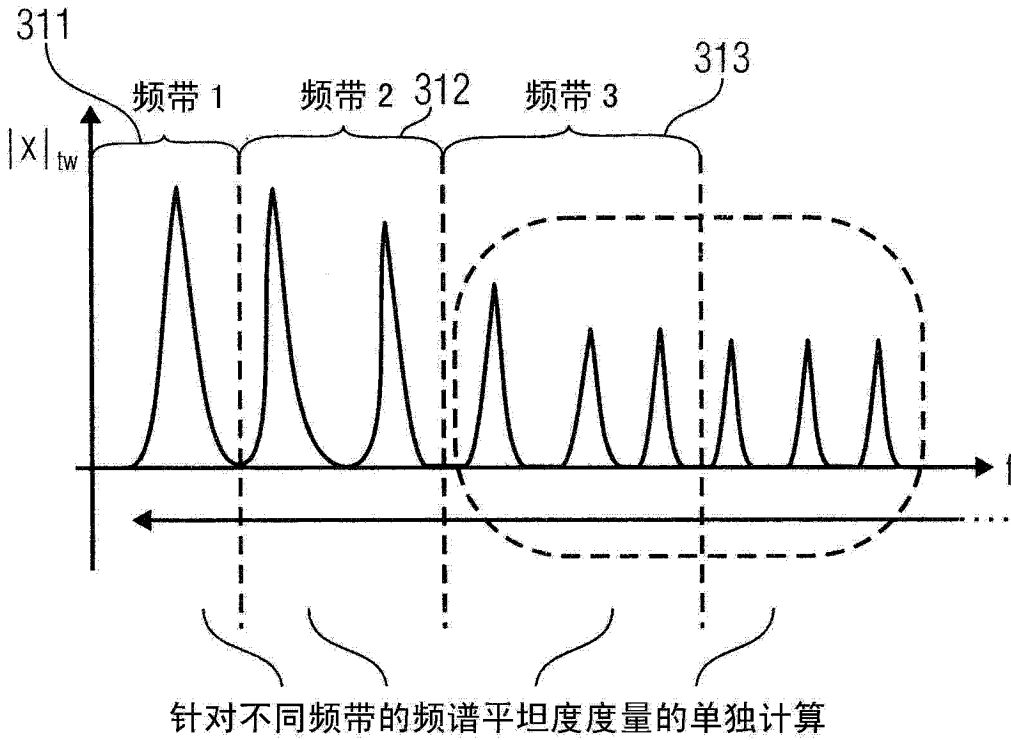


图 3C

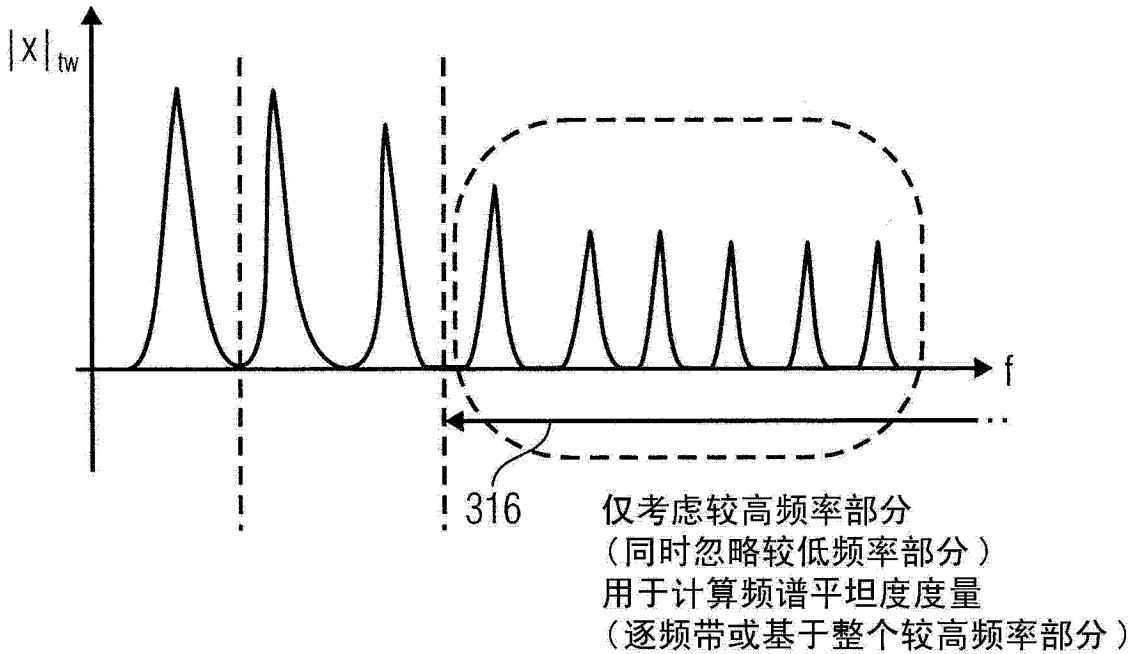


图 3D

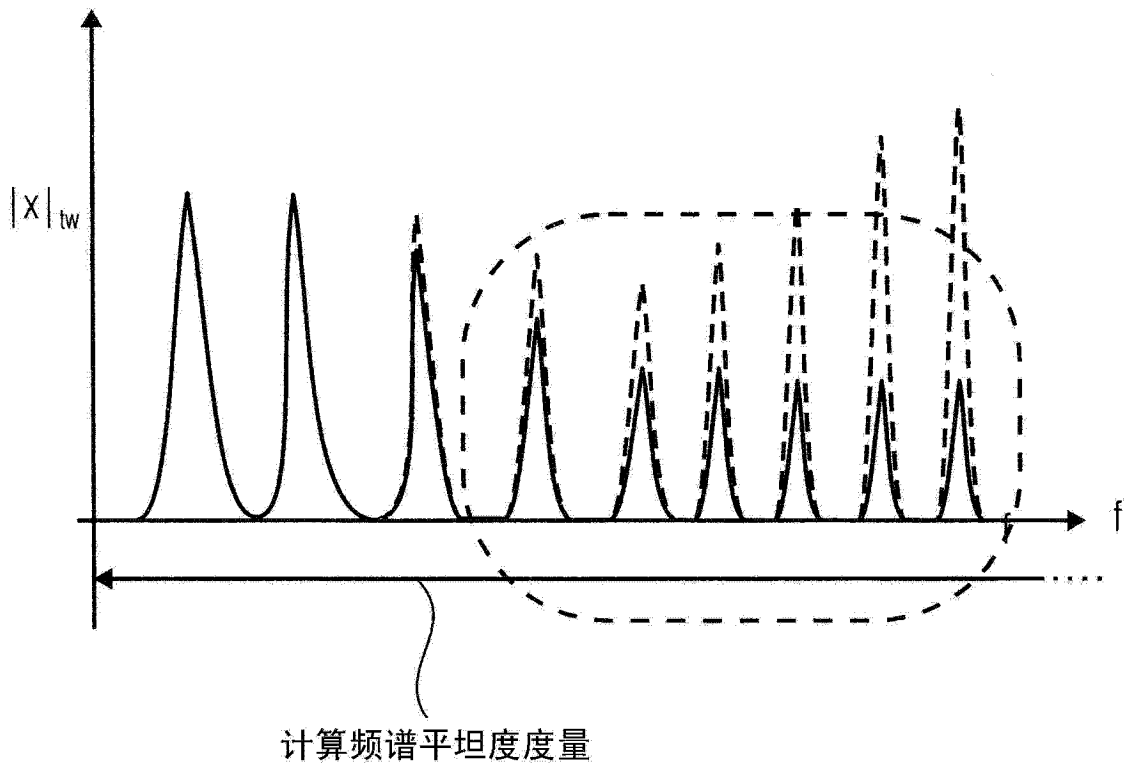


图 3E

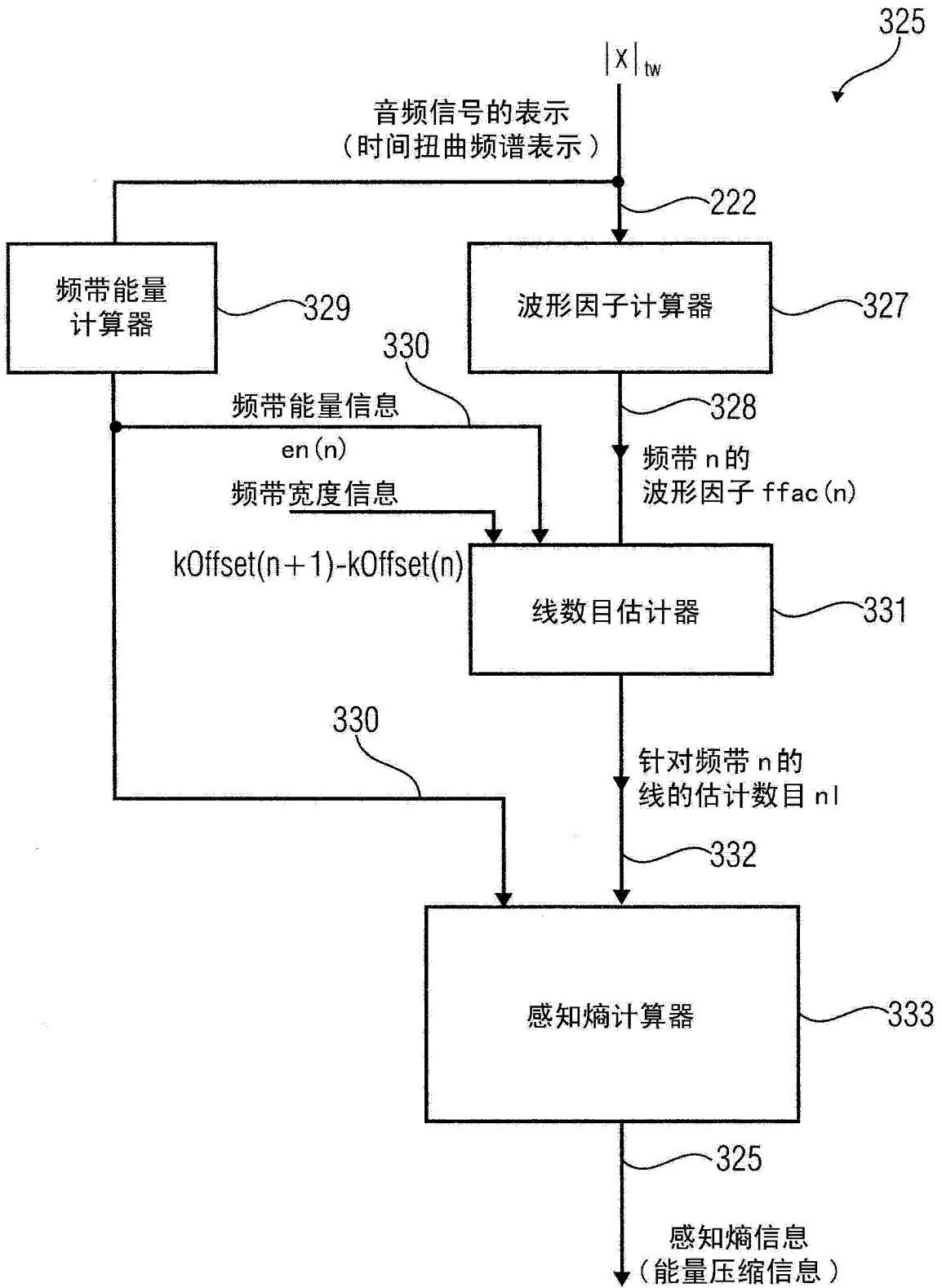


图 3F

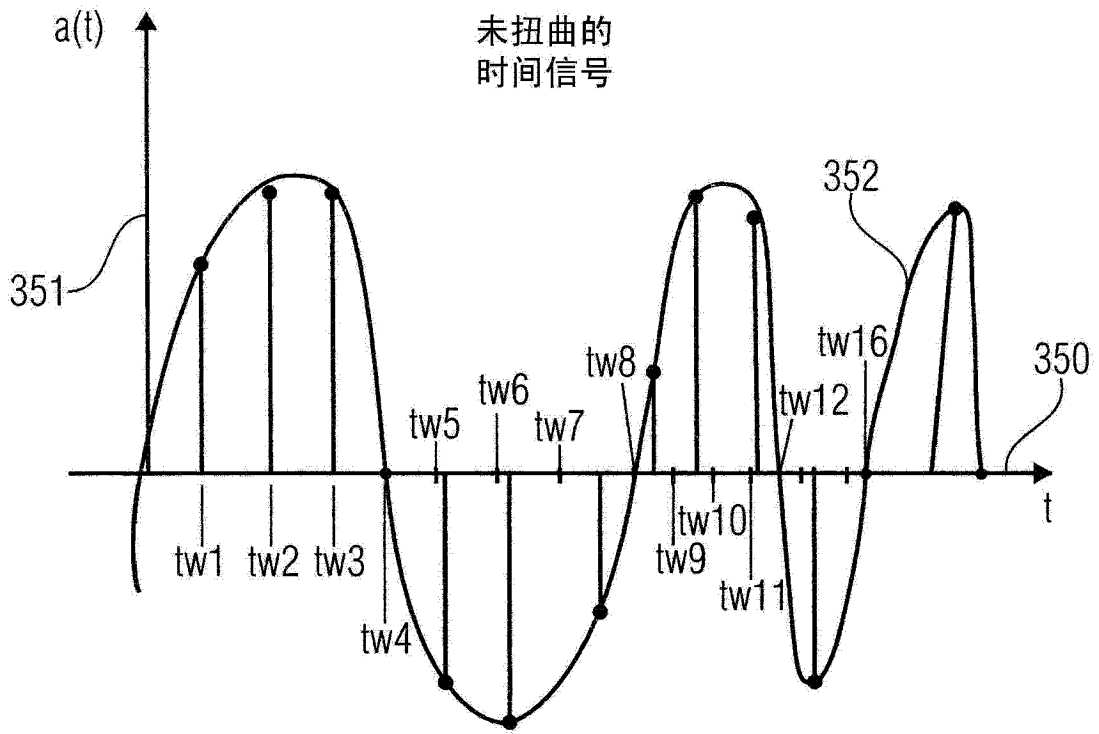


图 3G

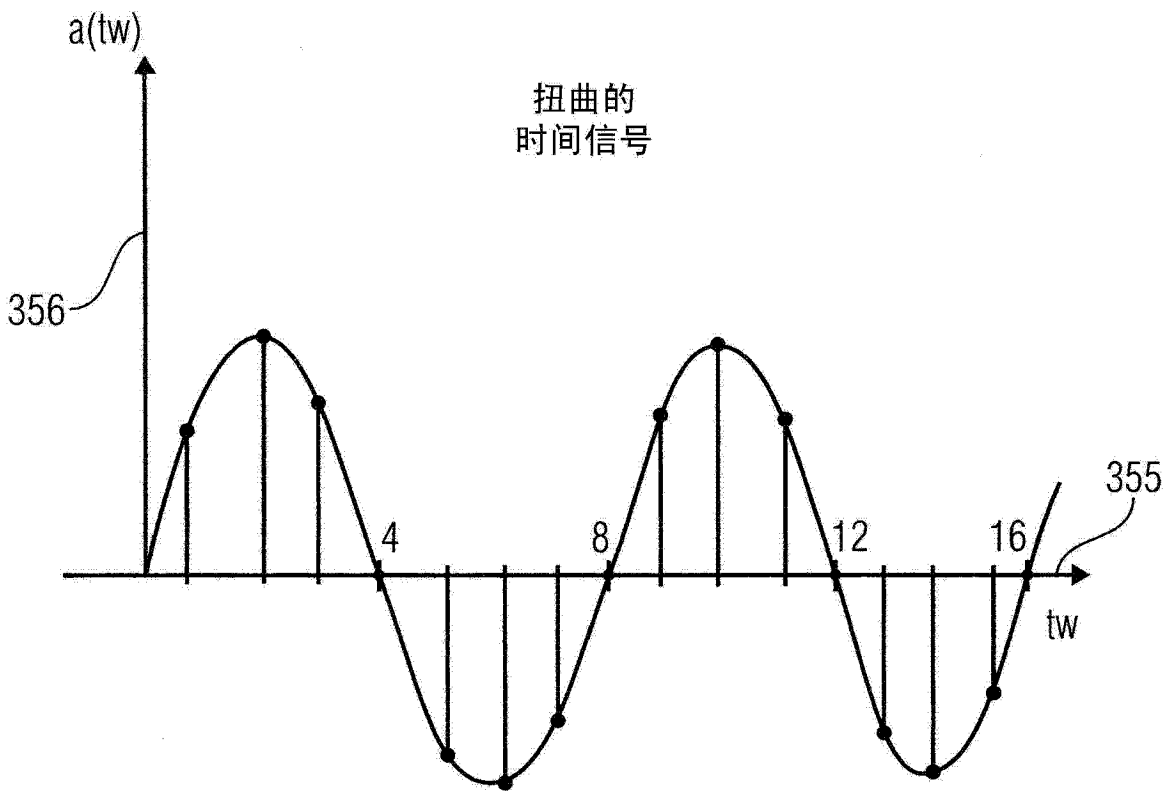


图 3H

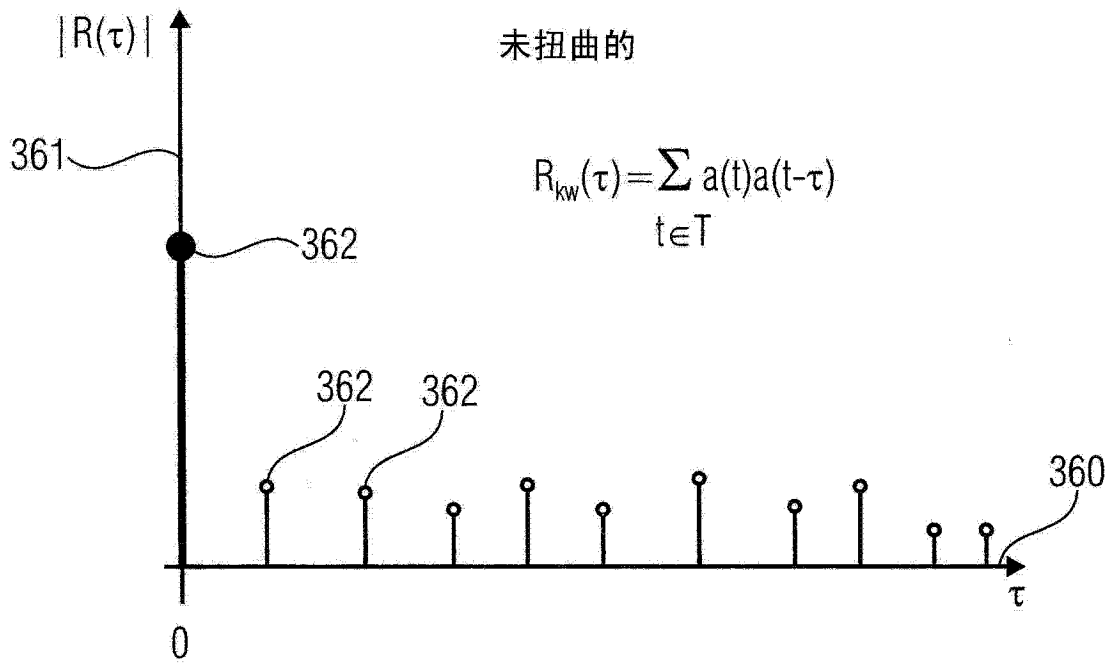


图 3I

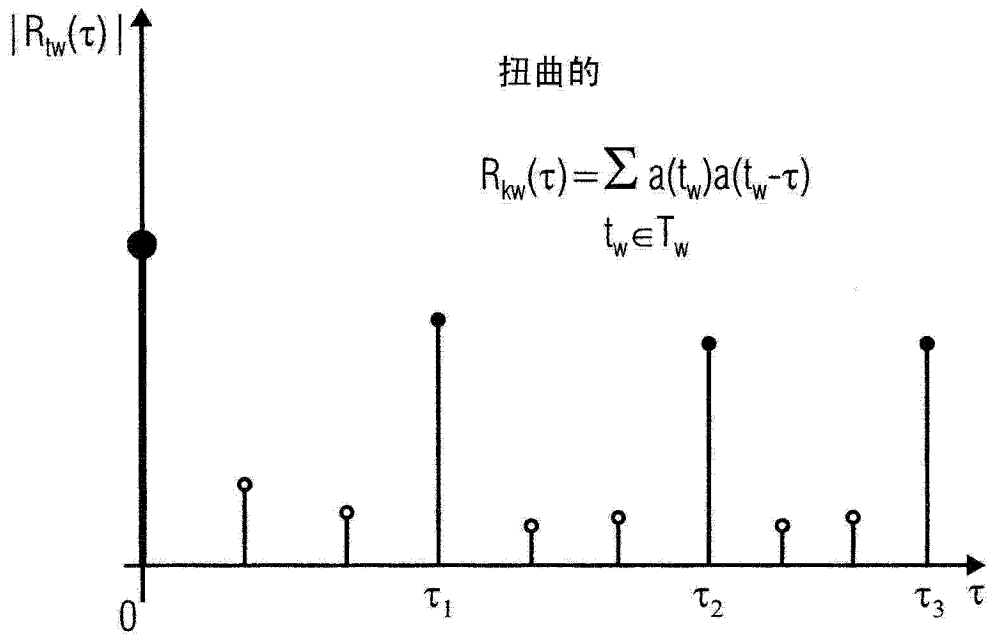


图 3J

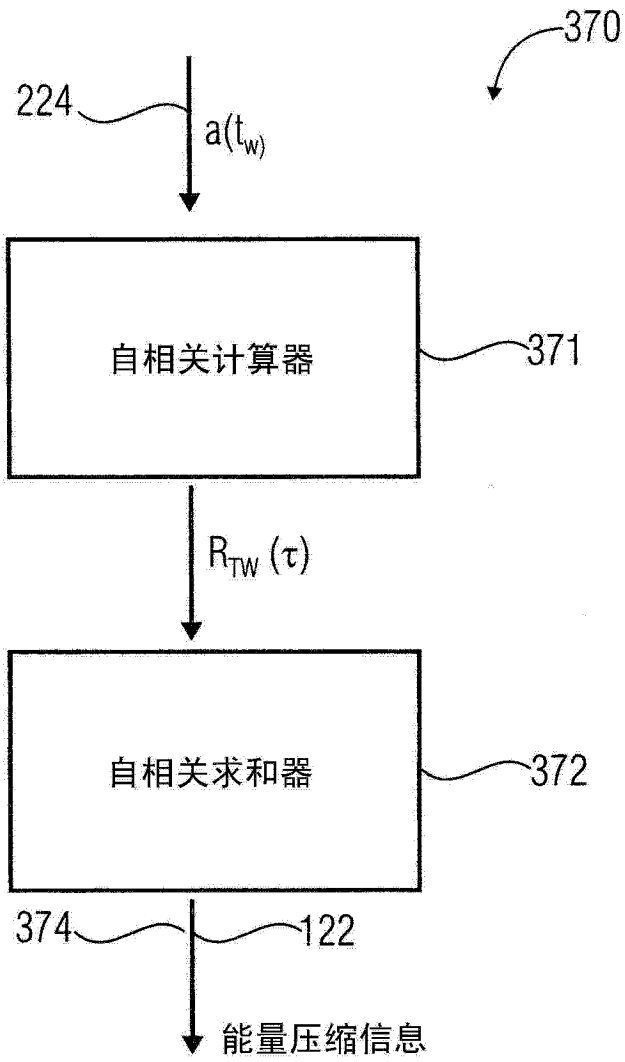


图 3K

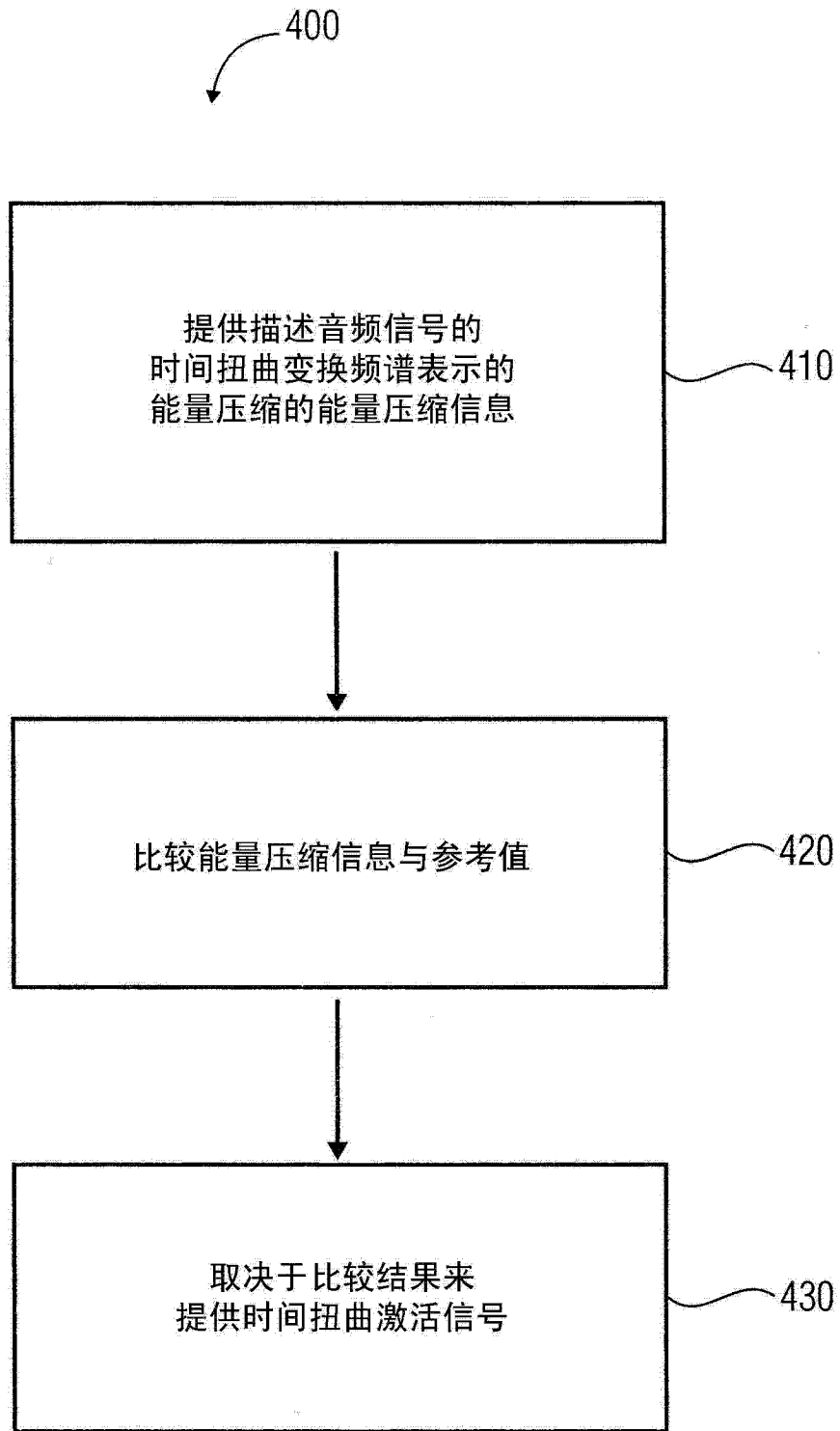


图 4A

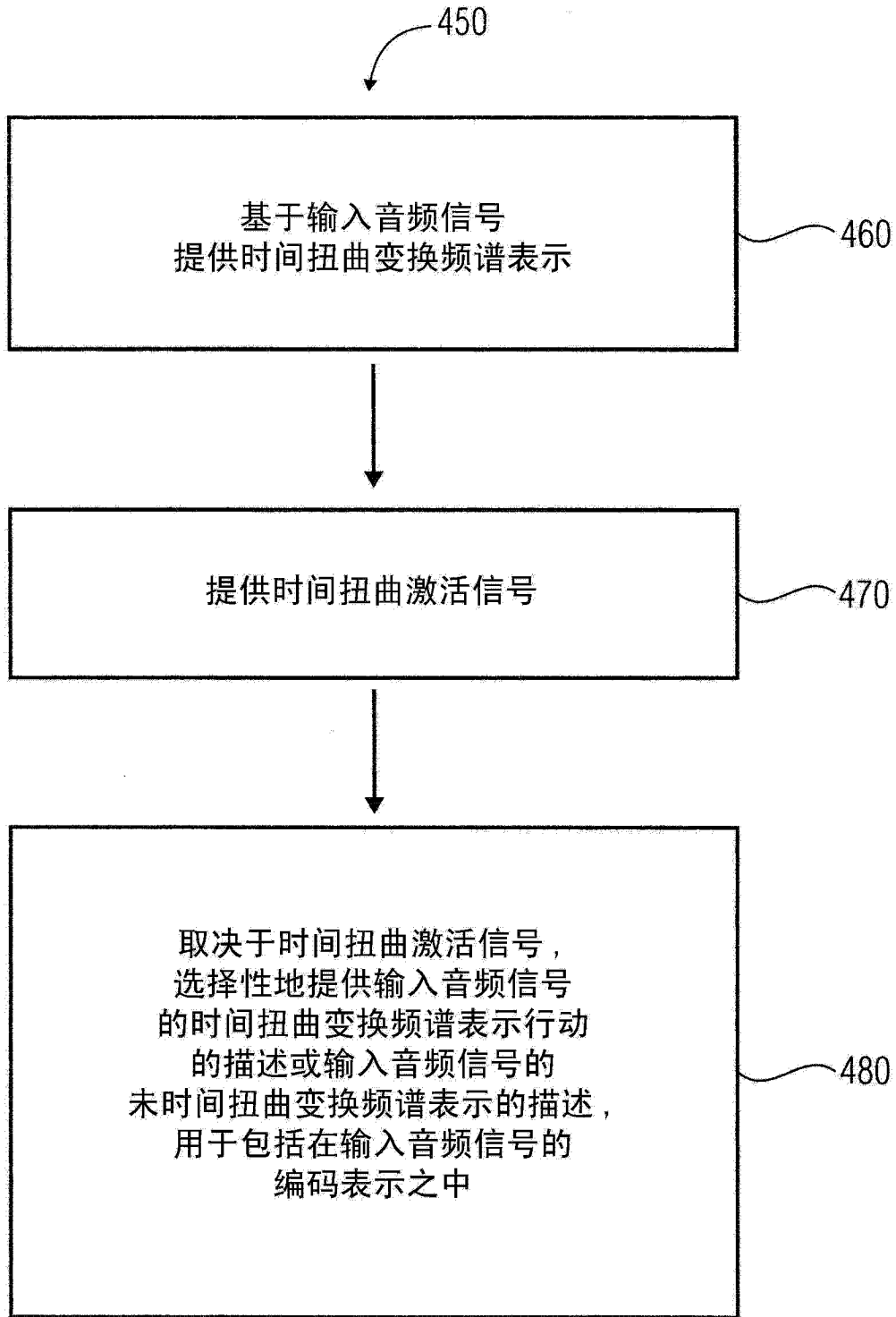


图 4B

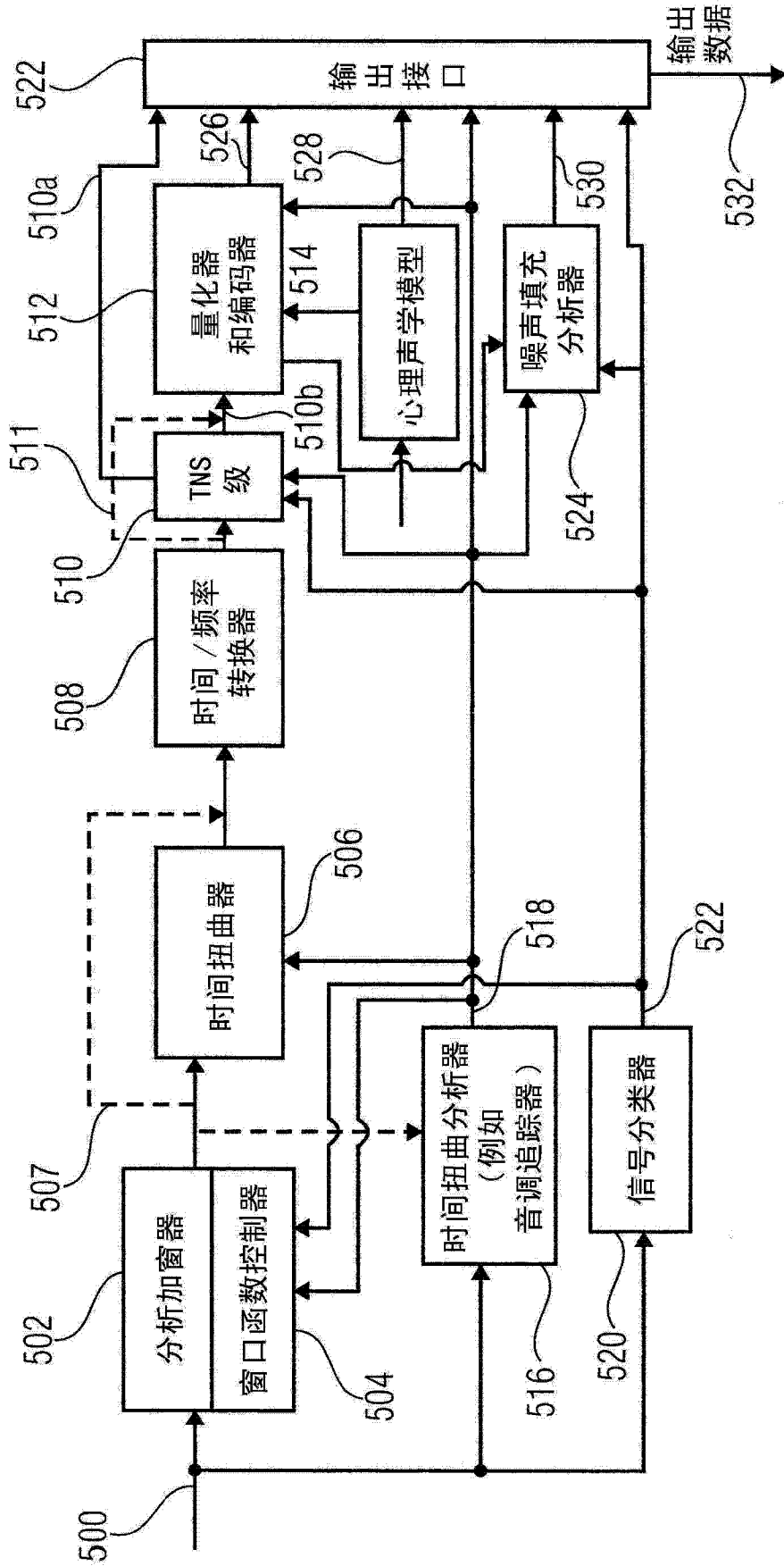


图 5A(编码器)

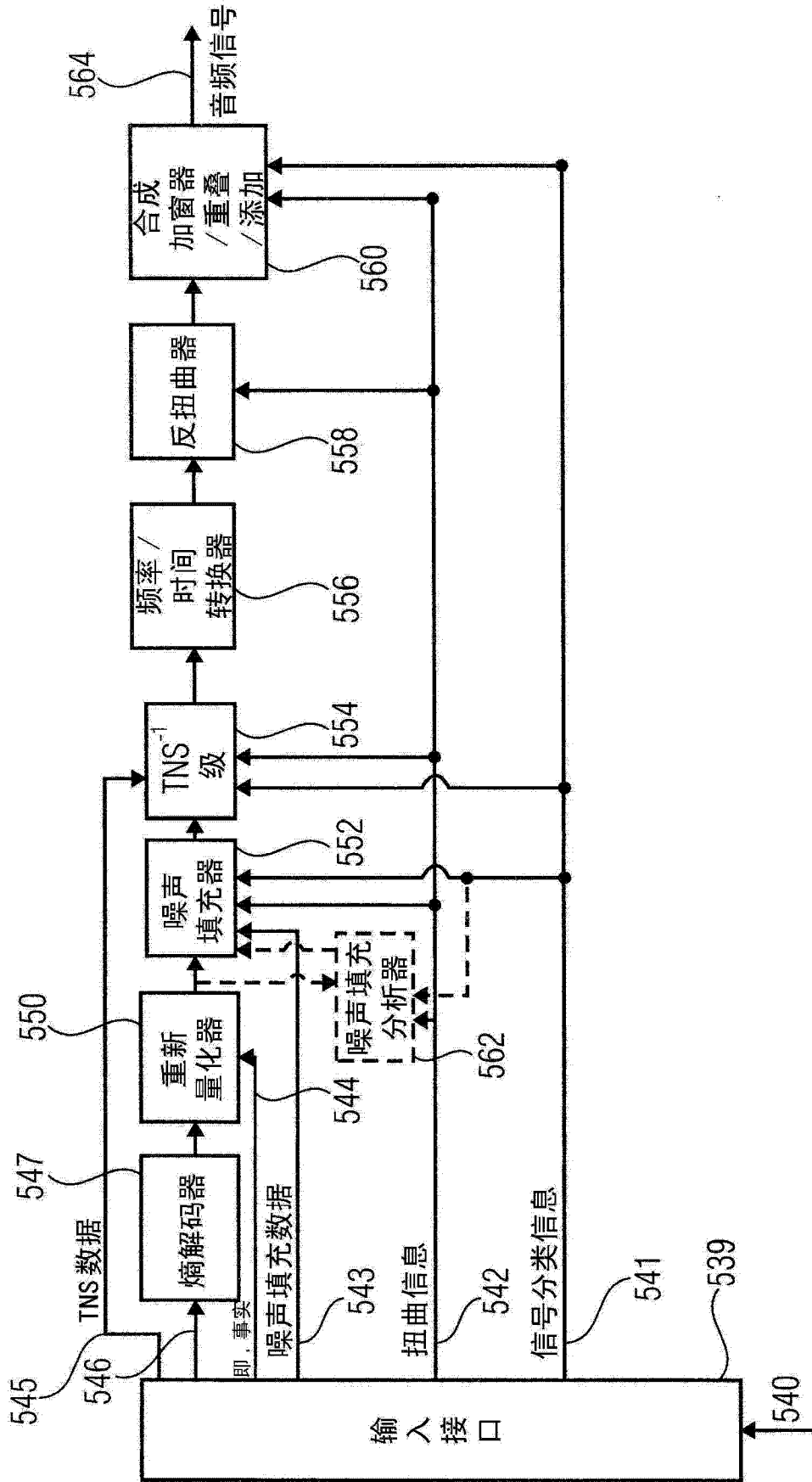


图 5B(解码器)

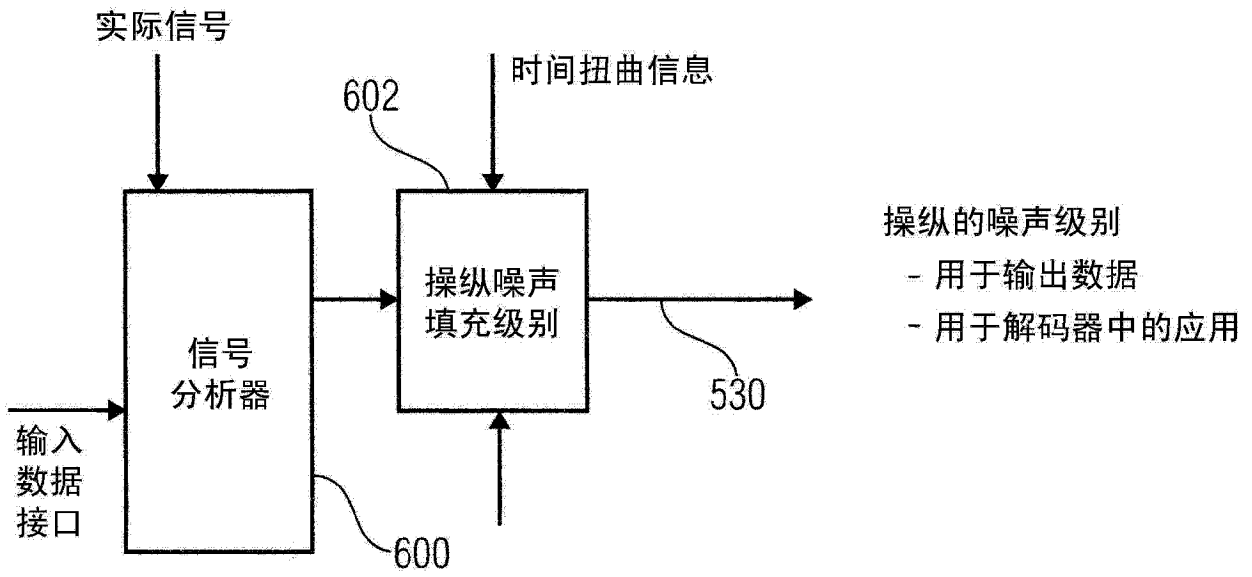


图 6A

语音分类		时间扭曲		信号分类		噪声填充级别
V	UV	是	否	语音	无语音	
			X		X	正常
		X			X	低
X		X		X		非常低/零
	X		X	X		正常

图 6B

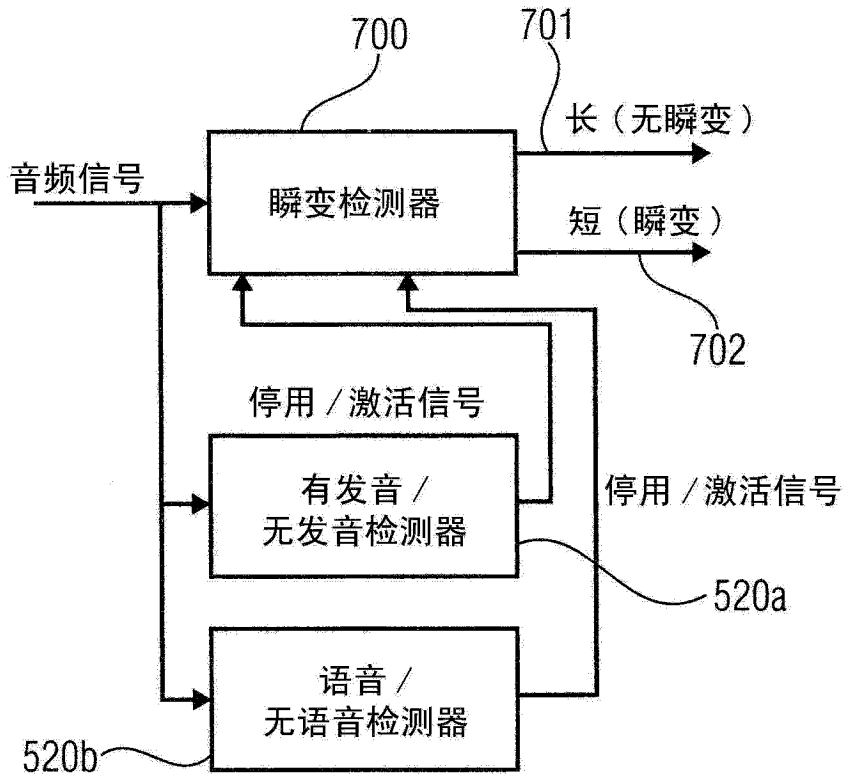


图 7A

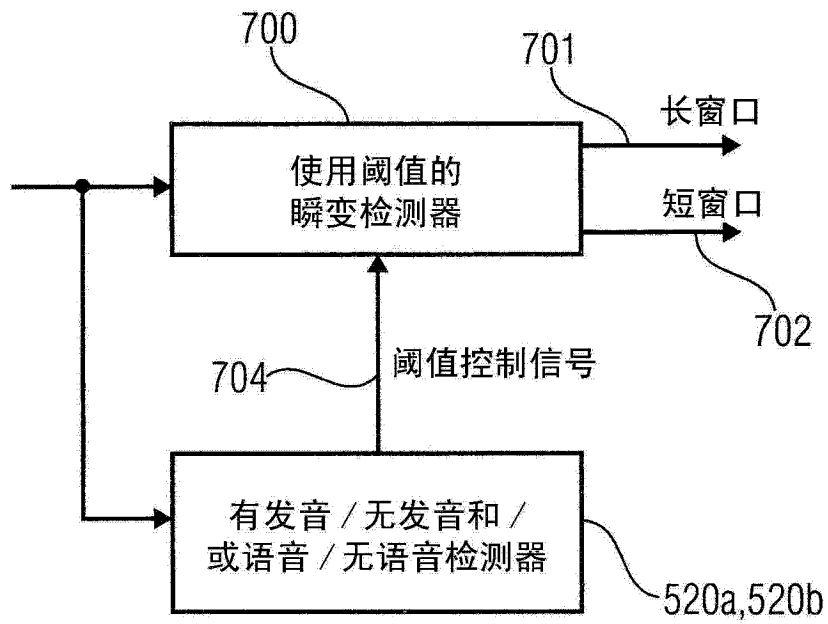


图 7B

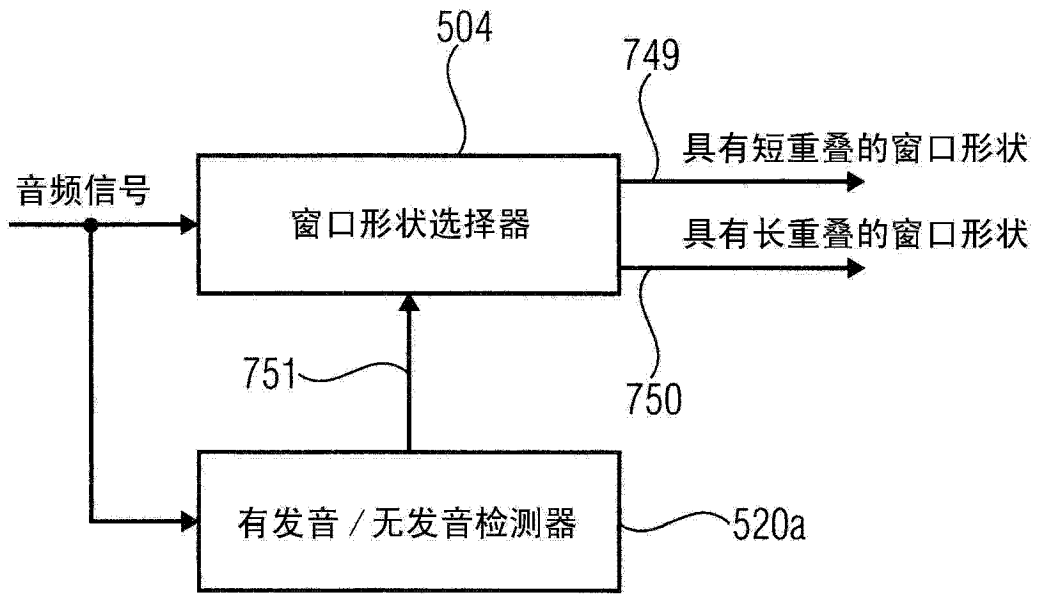


图 7C

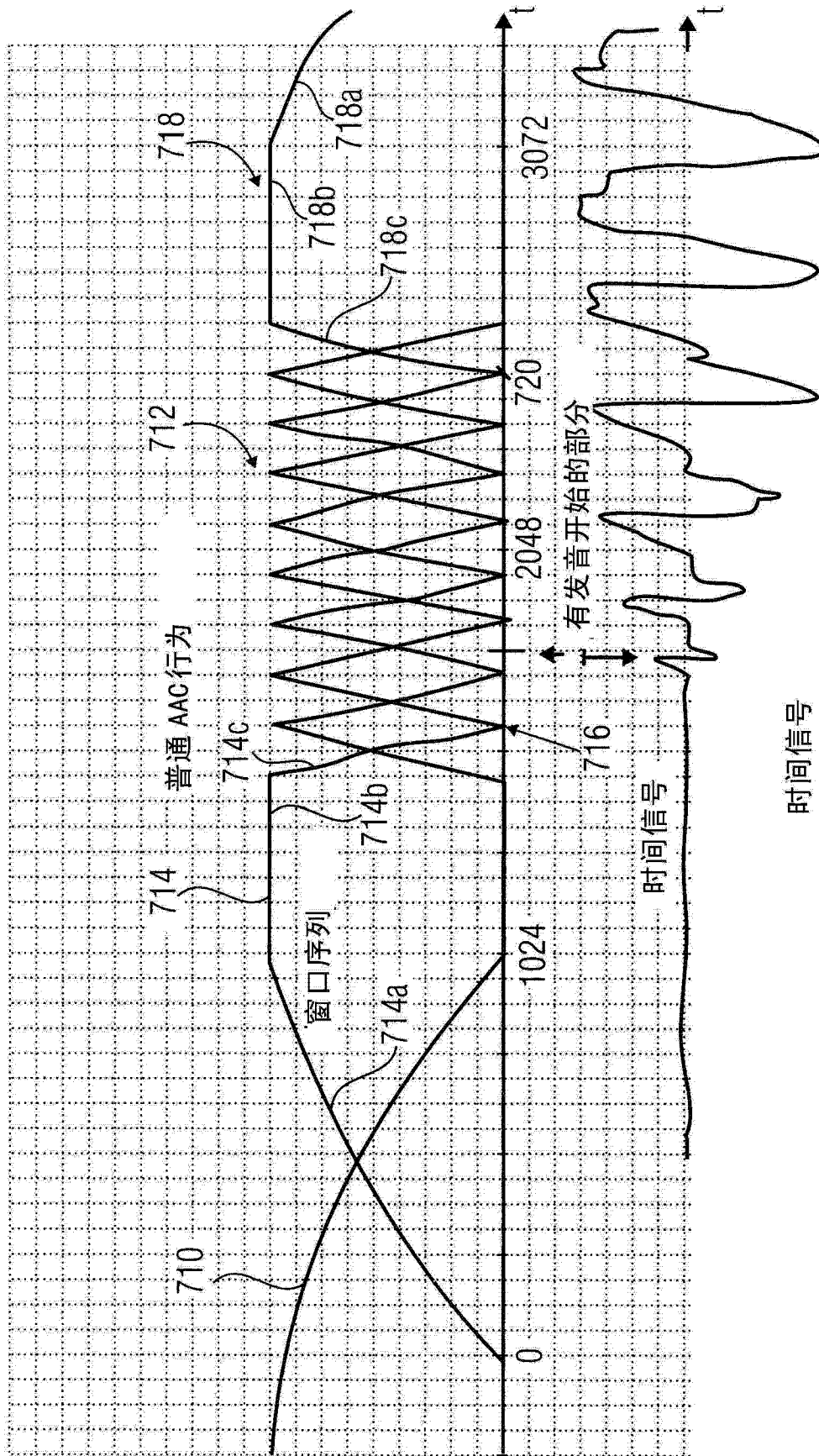


图 7D

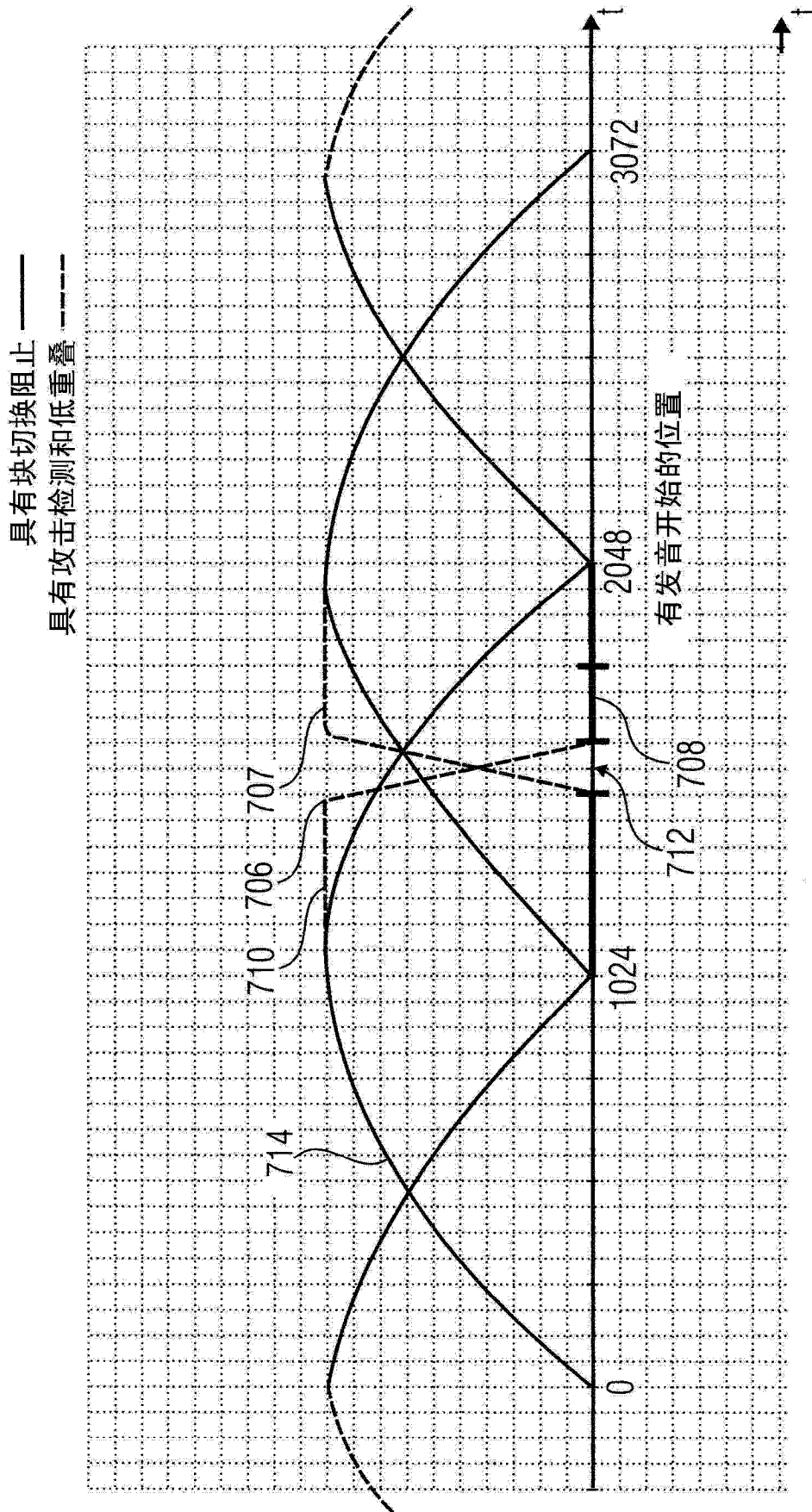


图 7E

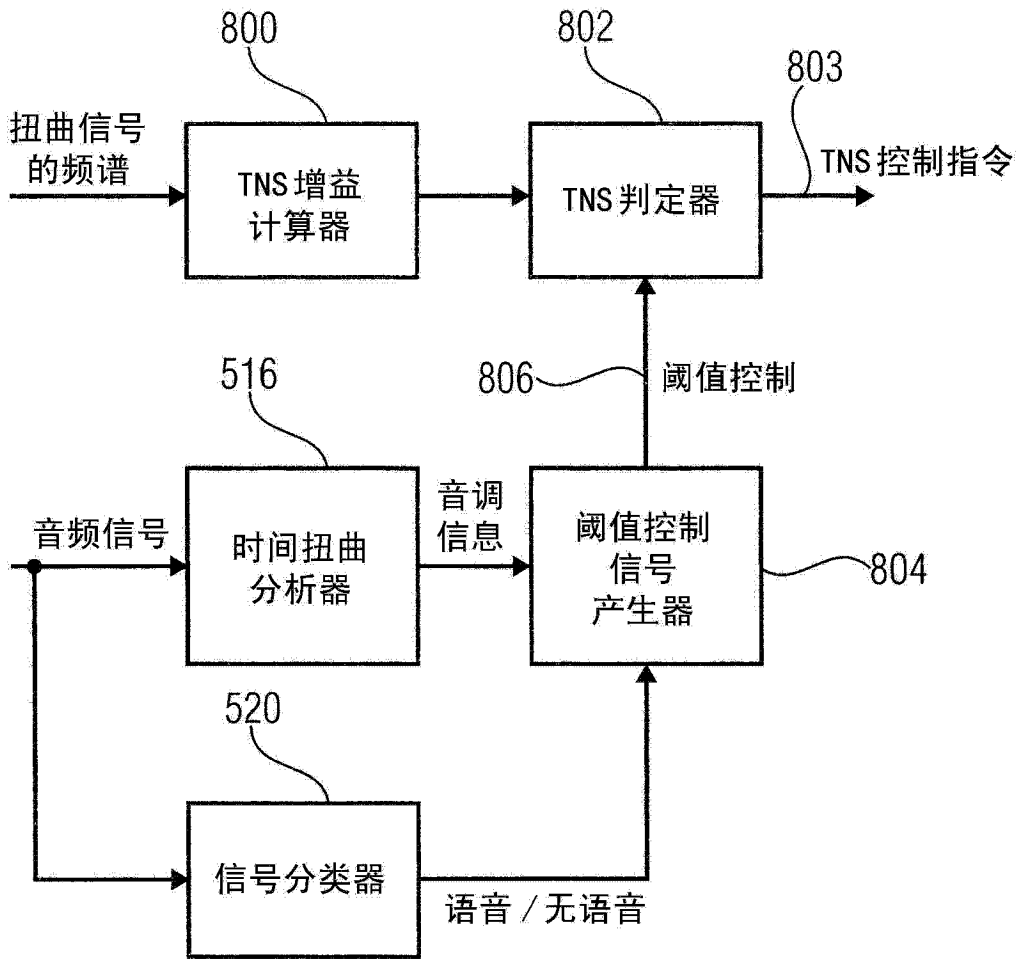
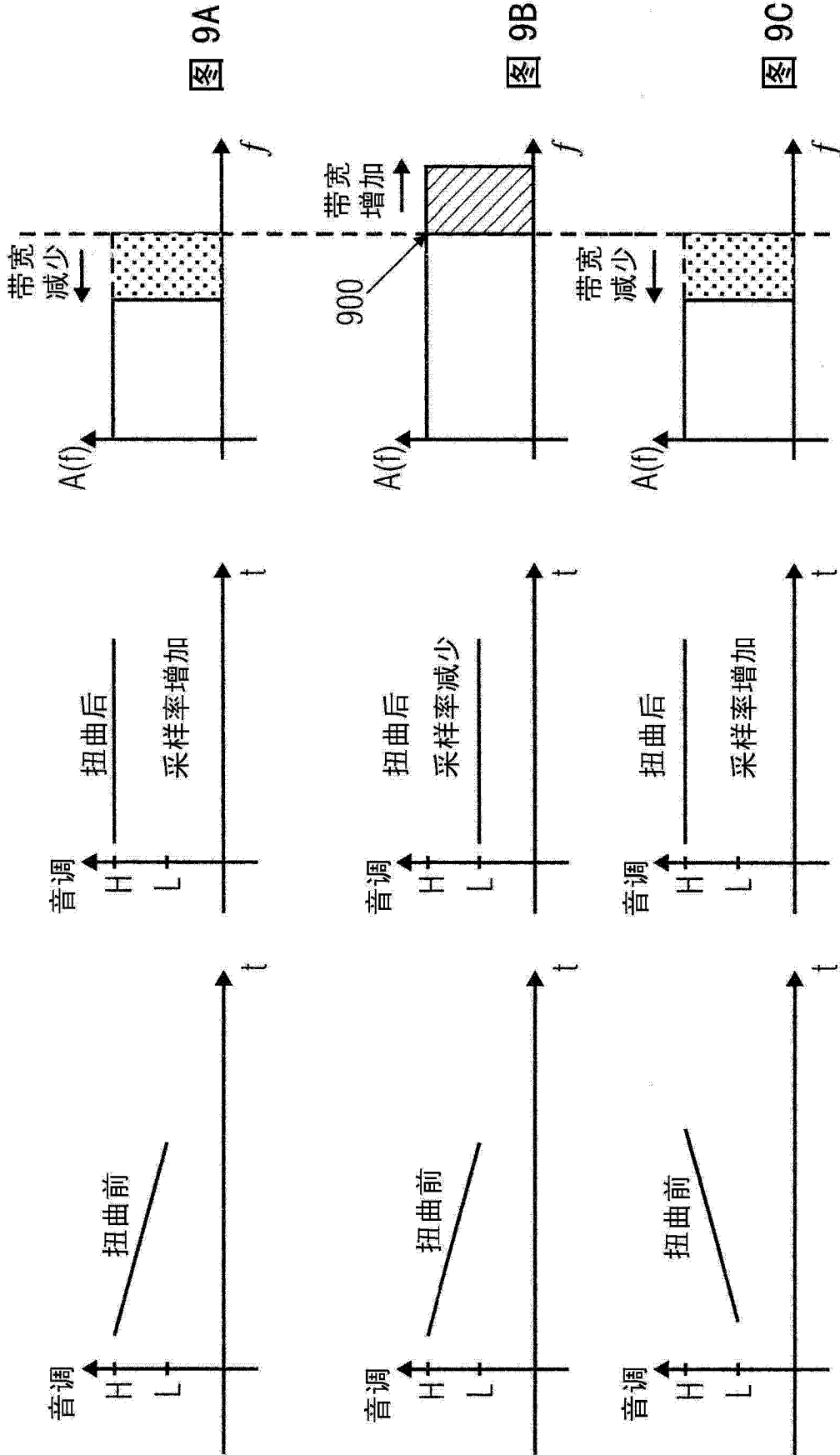


图 8A

音调控制		有发音	无发音	TNS判定阈值
		信号分类		
是	否	语音	无语音	
	X		X	正常
X			X	更低
X		X		(甚至)更低

图 8B



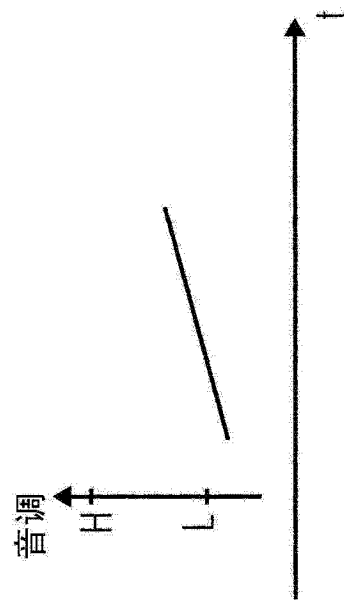
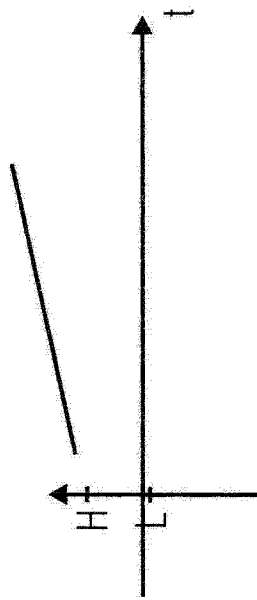
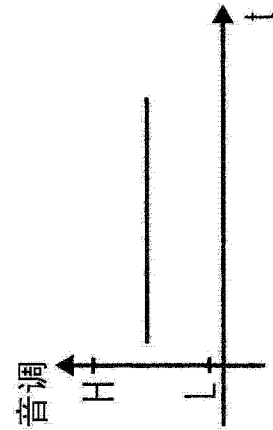
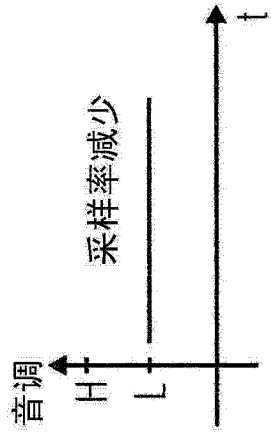
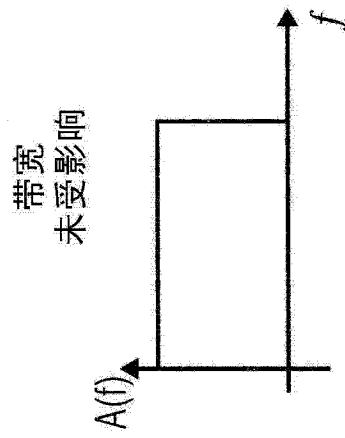
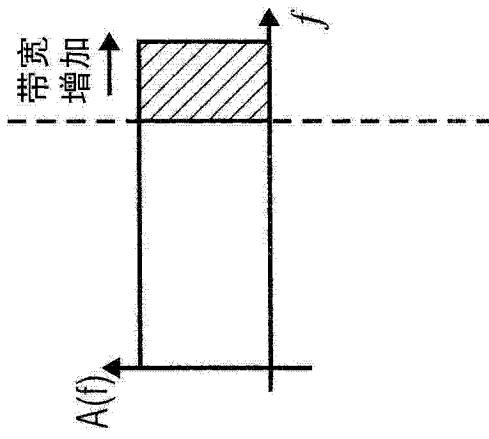


图 9D

图 9E

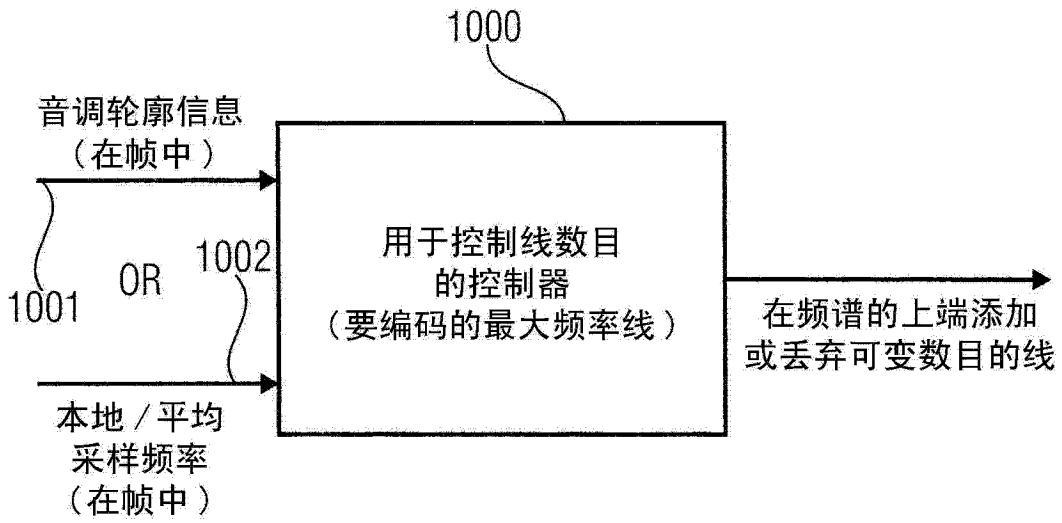


图 10A

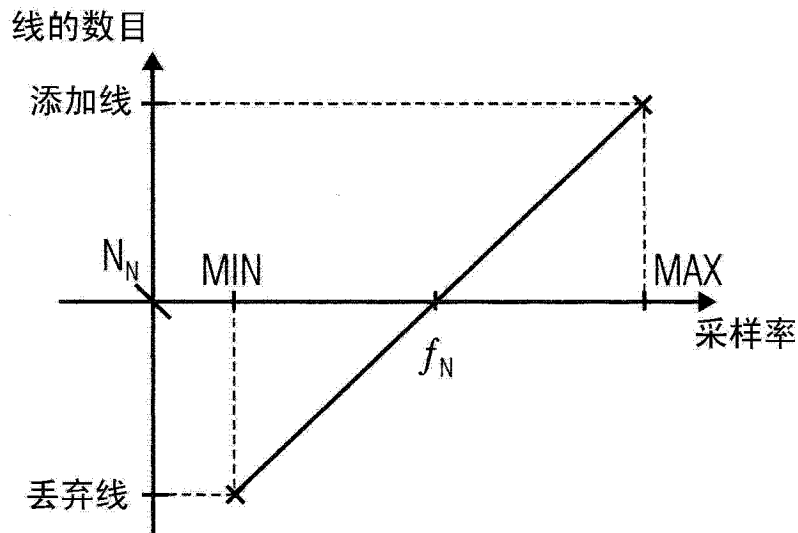


图 10B

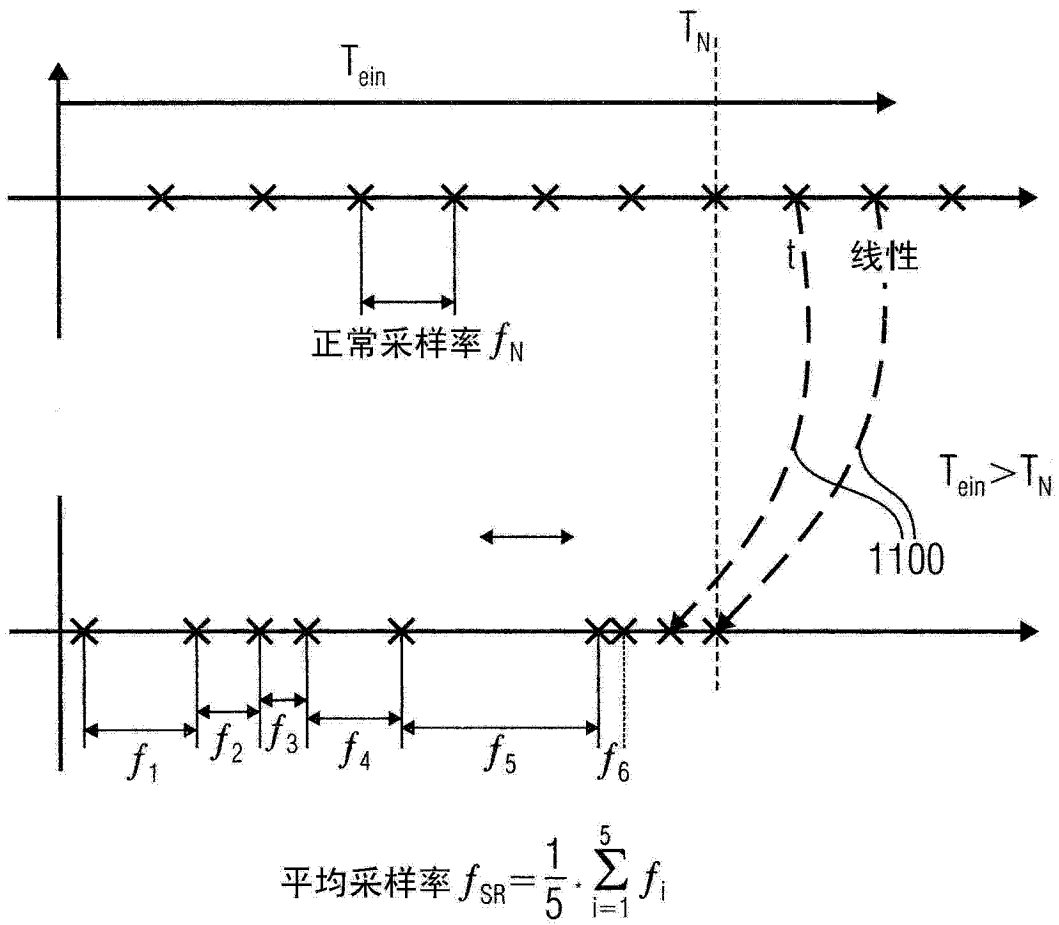


图 11

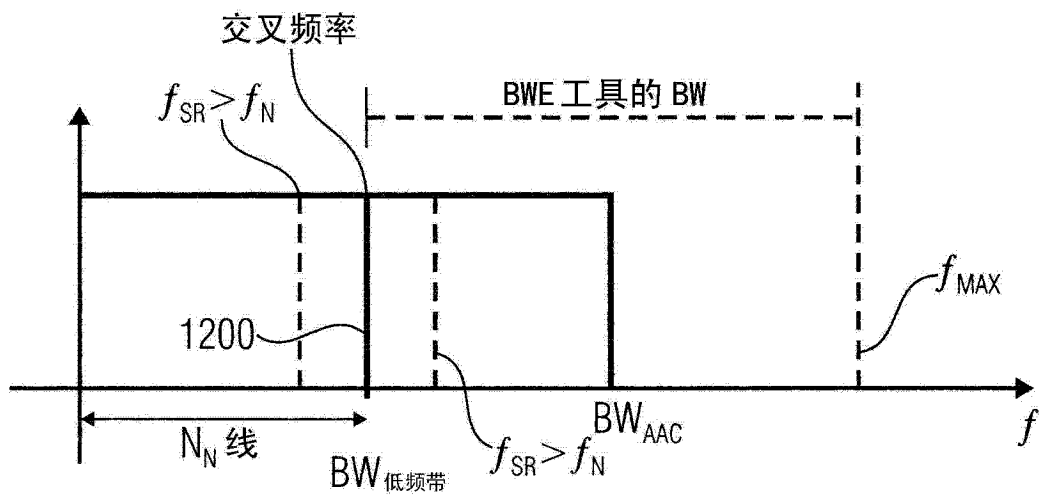


图 12A

f_{SR} : 本地采样率
 Δf : 频谱线的宽度 / 频谱线之间的距离
 N : 如处理器所确定的线的数目
 T_{ein} : 未扭曲帧的长度 (取决于扭曲特性)

T_{ein}	f_{SR}	Δf	N
$= T_N$	$= f_N$	$= \Delta f_N$	$= N_N$
$> T_N$	$> f_N$	$> \Delta f_N$	$< N_N$ 删除线
$< T_N$	$< f_N$	$< \Delta f_N$	$> N_N$ 添加线

图 12B