(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 111192595 B
(45)授权公告日 2023.09.22

(21)申请号 202010186693.3
(22)申请日 2015.05.12
(65)同一申请的已公布的文献号
申请公布号 CN 111192595 A
(43)申请公布日 2020.05.22
(30)优先权数据
61/993,639 2014.05.15 US
(62)分案原申请数据
201580026065.6 2015.05.12
(73)专利权人 瑞典爱立信有限公司
地址 瑞典斯德哥尔摩
(72)发明人 艾力克·诺维尔 斯蒂芬·布鲁恩
(74)专利代理机构 中科专利商标代理有限责任公司 11021
专利代理师 陈慧
(51)Int.Cl.
G10L 19/04(2013.01)
G10L 19/20(2013.01)
G10L 25/18(2013.01)
(56)对比文件
US 2010070269 A1,2010.03.18
CN 1325574 A,2001.12.05
CN 101681627 A,2010.03.24
CN 106415717 A,2017.02.15
CN 1659625 A,2005.08.24
CN 1930607 A,2007.03.14
审查员 李文娟

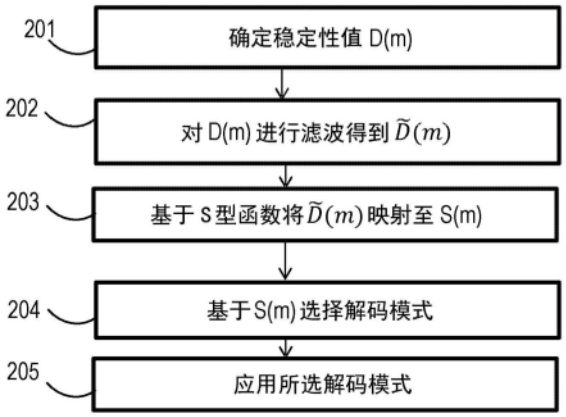
权利要求书2页 说明书18页 附图9页

(54)发明名称

音频信号分类和编码

(57)摘要

本发明涉及编解码器和信号分类器以及其中的基于音频信号特征进行的信号分类和编码模式选择的方法。一种由解码器执行的方法实施例包括,针对帧m:基于变换域中帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差,确定稳定性值D(m)。每个这种范围包括与音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。所述方法还包括:基于所述稳定性值D(m)从多个解码模式中选择解码模式;以及应用所选的解码模式。



CN 111192595 B

1.一种用于音频信号分类的方法,所述方法包括:

基于变换域中帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差,确定稳定性值D(m),每个范围包括与音频信号的分段的频谱带中的能量相关的频谱包络值的集合;

对所述稳定性值D(m)进行低通滤波,从而获得滤波后的稳定性值$\widetilde{D}(m)$;

通过使用S型函数将滤波后的稳定性值$\widetilde{D}(m)$映射至标量范围[0,1],从而获得稳定性参数S(m);以及

基于所述稳定性参数S(m)对所述音频信号进行分类。

2.根据权利要求1所述的方法,其中,对所述音频信号的分类包括:确定在帧m中表示的音频信号的分段是包括语音还是音乐。

3.根据权利要求1所述的方法,其中,对所述音频信号的分类还基于定义与所述音频信号中的语音和音乐之间的转变相关的状态转移概率的马尔可夫模型。

4.根据权利要求1所述的方法,其中,对所述音频信号的分类还基于指示帧m的频谱内容的瞬态结构的瞬态测量。

5.根据权利要求1所述的方法,其中,将所述稳定性值D(m)确定为

$$D(m)=\sqrt{\frac{1}{b_{end}-b_{start}+1}\sum_{b=b_{start}}^{b_{end}}\left(E(m,b)-E(m-1,b)\right)^2}$$

其中,b_i表示帧m中的频谱带,并且E(m,b)表示帧m中的频带b的能量测量。

6.一种音频信号分类器,所述分类器被配置为:

基于变换域中帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差,确定稳定性值D(m),每个范围包括与音频信号的分段的频谱带中的能量相关的频谱包络值的集合;

对所述稳定性值D(m)进行低通滤波,从而获得滤波后的稳定性值$\widetilde{D}(m)$;

通过使用S型函数将滤波后的稳定性值$\widetilde{D}(m)$映射至标量范围[0,1],从而获得稳定性参数S(m);以及

基于所述稳定性参数S(m)对所述音频信号进行分类。

7.根据权利要求6所述的分类器,其中,被配置为对所述音频信号进行分类的所述分类器包括:被配置为确定在帧m中表示的音频信号的分段是包括语音还是音乐的分类器。

8.根据权利要求6所述的分类器,其中,被配置为对所述音频信号进行分类的所述分类器还被配置为:基于定义与所述音频信号中的语音和音乐之间的转变相关的状态转移概率的马尔可夫模型,对所述音频信号进行分类。

9.根据权利要求6所述的分类器,被配置为对所述音频信号进行分类的所述分类器还被配置为:基于指示帧m的频谱内容的瞬态结构的瞬态测量对所述音频信号进行分类。

10.根据权利要求6所述的分类器,其中,将所述稳定性值D(m)确定为

$$D(m)=\sqrt{\frac{1}{b_{end}-b_{start}+1}\sum_{b=b_{start}}^{b_{end}}\left(E(m,b)-E(m-1,b)\right)^2}$$

其中，b_i表示帧m中的频谱带，并且E(m,b)表示帧m中的频带b的能量测量。

11.一种主机设备，所述主机设备包括根据权利要求6-10中任一项所述的音频信号分类器。

12.根据权利要求11所述的主机设备，所述主机设备被配置为基于由所述信号分类器执行的分类结果来从用于错误隐藏的多种方法中选择出一种用于错误隐藏的方法。

13.一种音频编码器，所述音频编码器包括根据权利要求6-10中任一项所述的音频信号分类器。

14.一种音频解码器，所述音频解码器包括根据权利要求6-10中任一项所述的音频信号分类器。

音频信号分类和编码

[0001] 本申请是申请日为2015年5月12日、申请号为201580026065.6的中国发明专利申请“音频信号分类和编码”的分案申请。

技术领域

[0002] 本发明涉及音频编码，并且更具体地涉及为了编码对输入信号特征进行分析和匹配。

背景技术

[0003] 蜂窝通信网络朝着更高的数据速率、改进的容量和改进的覆盖演进。在第三代合作伙伴项目(3GPP)标准主体中，已经开发了并且当前也正在开发若干技术。

[0004] LTE(长期演进)是标准化技术的示例。在LTE中，针对下行链路使用基于OFDM(正交频分复用)的接入技术，而针对上行链路使用基于单载波FDMA(SC-FDMA)的接入技术。通常在考虑每个无线终端的瞬时业务模式和无线电传播特征的情况下通过使用快速调度自适应地执行在下行链路和上行链路二者上对无线终端(也被称为用户设备，UE)的资源分配。LTE上的一种数据类型是例如用于语音会话或流音频的音频数据。

[0005] 为了改进低比特率语音和音频编码的性能，利用关于信号特征的先验知识并采用信号建模，这是公知的。在使用更复杂信号的情况下，可以针对信号的不同部分使用若干编码模型或编码模式。这些编码模式还可能涉及用于处理信道错误和丢失分组的不同策略。在任何时候选择适当的编码模式是有益的。

发明内容

[0006] 本文所述的解决方案涉及可用于编码方法选择和/或错误隐藏方法选择二者(其在本文中已被概括为编码模式的选择)的信号分类或区分的低复杂度的、稳定的适配。在错误隐藏的情况下，该解决方案涉及解码器。

[0007] 根据第一方案，提供了一种对音频信号进行解码的方法。所述方法包括，针对帧m：基于变换域中帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差，确定稳定性值D(m)。每个这种范围包括与音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。所述方法还包括：基于所述稳定性值D(m)从多个编码模式中选择编码模式；以及应用所选的解码模式。

[0008] 根据第二方案，提供了一种用于对音频信号进行解码的解码器。所述解码器被配置为，针对帧m：基于变换域中帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差，确定稳定性值D(m)。每个这种范围包括与音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。所述编码器还被配置为：基于所述稳定性值D(m)从多个解码模式中选择编码模式；以及应用所选的解码模式。

[0009] 根据第三方案，提供了一种对音频信号进行编码的方法。所述方法包括，针对帧m：基于变换域中帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差，确定稳

定性值D(m)。每个这种范围包括与音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。所述方法还包括:基于所述稳定性值D(m)从多个解码模式中选择解码模式;以及应用所选的编码模式。

[0010] 根据第四方案,提供了一种用于对音频信号进行编码的编码器。所述编码器被配置为,针对帧m:基于变换域中帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差,确定稳定性值D(m)。每个这种范围包括与音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。所述编码器还被配置为:基于所述稳定性值D(m)从多个解码模式中选择解码模式;以及应用所选的编码模式。

[0011] 根据第五方案,提供了一种音频信号分类方法。所述方法包括,针对语音信号的帧m:基于变换域中帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差,确定稳定性值D(m),每个范围包括与所述音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。所述方法还包括:基于稳定性值D(m)对所述语音信号分类。

[0012] 根据第六方案,提供了一种音频信号分类器。音频信号被配置为,针对语音信号的帧m:基于变换域中帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差,确定稳定性值D(m),每个范围包括与所述音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合;以及还基于稳定性值D(m)对所述语音信号分类。

[0013] 根据第七方案,提供了一种主机设备,包括根据第二方案的解码器。

[0014] 根据第八方案,提供了一种主机设备,包括根据第四方案的编码器。

[0015] 根据第九方案,提供了一种主机设备,包括根据第六方案的信号分类器。

[0016] 根据第十方案,提供了一种计算机程序,其包括指令,所述指令当在至少一个处理器上执行时使所述至少一个处理器执行根据第一、第三和/或第六方案所述的方法。

[0017] 根据第十一方案,提供了一种包含第九方案所述的计算机程序的载体,其中,所述载体是电信号、光信号、无线电信号或计算机可读存储介质中的一种。

附图说明

[0018] 现在将参照附图通过示例描述本发明,在附图中:

[0019] 图1是示出了可以应用本文介绍介绍的实施例的蜂窝网的示意图。

[0020] 图2a和图2b是示出了根据示例性实施例的由解码器执行的方法的流程图。

[0021] 图3a是示出从滤波后的稳定性值到稳定性参数的映射曲线的示意图;

[0022] 图3b是示出从滤波后的稳定性值到稳定性参数的映射曲线的示意图,其中映射曲线是从离散值获得的;

[0023] 图4是示出接收的音频帧的信号的频谱包络的示意图;

[0024] 图5a-b是示出在主机设备中执行的用于选择分组丢失隐藏过程的方法的流程图;

[0025] 图6a-c是根据示例实施例示出解码器的不同实施方式的示意框图。

[0026] 图7a-c是根据示例实施例示出编码器的不同实施方式的示意框图。

[0027] 图8a-c是根据示例实施例示出分类器的不同实施方式的示意框图。

[0028] 图9是示出了无线终端的一些组件的示意图;

[0029] 图10是示出了转码节点的一些组件的示意图;以及

[0030] 图11示出了包括计算机可读装置的计算机程序产品的一个示例。

具体实施方式

[0031] 现将在下文参考示出本发明的特定实施例的附图来更全面地描述本发明。然而，本发明可以用多种不同形式来实施，并且不应当被解释为受到本文阐述的实施例的限制；相反，这些实施例是作为示例给出的，使得本公开将是透彻和完整的，并将向本领域技术人员充分地传达本发明的范围。在本说明书全文中，相似的标记指相似的元素。

[0032] 图1是示出了可以应用本文介绍介绍的实施例的蜂窝网8的示意图。蜂窝网8包括核心网3和一个或多个无线电基站1，这里所述无线电基站1具有演进节点B(也称为eNodeB或eNB)的形式的。无线电基站1也可以是节点B、BTS(基本收发机站)和/或BSS(基站子系统)等的形式。无线电基站1提供与多个无线终端2的无线电连接。术语无线终端也称为移动通信终端、用户设备(UE)、移动终端、用户终端、用户代理、无线设备、机器对机器设备等，并且可以是例如今天通常被称为移动电话或具有无线连接或固定安装端的平板/膝上型计算机。

[0033] 只要下文所描述的原理是适用的，蜂窝网8可以例如符合LTE(长期演进)、W-CDMA(宽带码分多址)、EDGE(增强型数据速率GSM(全球移动通信系统)演进)、GPRS(通用分组无线电服务)、CDMA2000(码分多址2000)或任何其它当前或未来的无线网络(诸如高级LTE)中的任意一个或其组合。

[0034] 在无线电接口上执行无线终端2与无线电基站1之间的来自无线终端2的上行链路(UL)4a通信和去往无线终端2的下行链路(DL)4b通信。由于衰落、多径传播、干扰等影响，到每个无线终端2的无线电接口的质量可能随时间并根据无线终端2的位置而变化。

[0035] 无线电基站1还与核心网3连接，核心网3用于与中心功能和外部网络7(例如公共交换电话网(PSTN)和/或因特网)连接。

[0036] 可以例如由无线终端2和转码节点5对音频数据进行编码和解码，转码节点5是被布置为执行音频的转码的网络节点。可以例如在MGW(媒体网关)、SBG(会话边界网关)/BGF(边界网关功能)或MRFP(媒体资源功能处理器)中实现转码节点5。因此，无线终端2和转码节点5二者均是包括相应音频编码器和解码器的主机设备。

[0037] 使用一组错误恢复或错误隐藏方法和根据瞬时信号特征选择适当的隐藏策略，在许多情况下可以提高重建的音频信号的质量。

[0038] 为了选择最佳编码/解码模式，编码器和/或解码器可以在分析合成(analysis-by-synthesis)(也称为闭环方式)中尝试所有可用模式，或者它可以依赖于基于信号分析模式对编码做出判定(也称为开环判定)的信号分类器。语音信号的典型信号类别是发声和不发声语音。对于一般的音频信号，通常在语音、音乐和潜在的背景噪声信号之间进行区分。类似的分类可用于控制错误恢复或错误隐藏方法。

[0039] 然而，信号分类器可能涉及在计算复杂性和存储器资源方面具有高成本的信号分析。为所有信号找到合适的分类也是困难的问题。

[0040] 可以通过使用编码或解码方法中已经可用的编解码器参数的信号分类方法的应用来避免计算复杂度的问题，从而增加非常少的附加计算复杂度。信号分类方法也可以根据目前的编码模式使用不同的参数，以便即使当编码模式改变时也给出可靠的控制参数。这给出了可以用于编码方法选择和错误隐藏方法选择二者的信号分类的低复杂度且稳定的适配。

[0041] 实施例可以应用于在频域或变换域中操作的音频编解码器。在编码器处，输入样本x(n)被划分为固定长度或变化长度的时间段或帧。为了表示帧m的样本，写入x(m,n)。通常，使用20ms的固定长度，具有以下选择：针对快速时间变化(例如在瞬态声音处)可以选择使用较短的窗口长度或帧长度。通过频率变换将输入采样变换至频域。由于修改的离散余弦变换的编码的适用性，许多音频编解码器采用修改的离散余弦变换(MDCT)。也可以使用其他变换，例如DCT(离散余弦变换)或DFT(离散傅立叶变换)。使用以下关系找到帧m的MDCT频谱系数：

[0042] $$X(m,k)=\sum_{k=0}^{2L-1}x(m,n)\cos\left(\frac{\pi}{L}+\frac{1}{2}+\frac{L}{2}\right)\left(k+\frac{1}{2}\right)$$

[0043] 其中X(m,k)表示帧m中的MDCT系数k。将MDCT频谱的系数划分为组或频带。这些频带在大小上通常是不均匀的，对于低频使用较窄的频带，对于较高频使用较宽的频带。这旨在模仿人类听觉感知的频率分辨率和有损编码方案的相关设计。频带b的系数是MDCT系数的矢量：

[0044] $X(m,k),k=k_{start(b)},k_{start(b)}+1,\ldots,k_{end(b)}$

[0045] 其中$k_{start(b)}$和$k_{end(b)}$表示带b的开始和结束索引。然后计算每个频带的能量或均方根(RMS)值

[0046] $$E(m,b)=\sqrt{\frac{1}{k_{start(b)}-k_{end(b)}+1}\sum_{k=k_{start(b)}}^{k_{end(b)}}X(m,k)^2}$$

[0047] 频带能量E(m,b)形成MDCT频谱的频谱粗糙结构或包络。使用合适的量化技术(例如使用差分编码结合熵编码)或矢量量化器(VQ)对其进行量化。量化步骤产生要存储或发送给解码器的量化索引，并还再现相应的量化包络值$\hat{E}(m,b)$。MDCT频谱用量化的频带能量进行归一化以形成归一化的MDCT频谱N(m,k)：

[0048] $$N(m,k)=\frac{1}{\hat{E}(m,b)}X(m,k),\quad k=k_{start(b)},k_{start(b)}+1,\ldots,k_{end(b)}$$

[0049] 使用合适的量化技术(例如结合差分编码和熵编码的标量量化器，或矢量量化技术)对归一化的MDCT频谱进行进一步量化。通常，量化涉及针对每个频带b产生比特分配R(b)，其被用于对每个频带进行编码。可以产生比特分配，其包括基于感知重要性将比特分配给各频带的感知模型。

[0050] 可能希望通过对信号特征的适应来进一步引导编码器和解码器处理。如果使用在编码器和解码器二者处均可用的量化参数来进行适配，则可以在编码器和解码器之间对适配进行同步，而不需要传输附加参数。

[0051] 这里描述的解决方案主要涉及使编码器和/或解码器处理适应于要编码或解码的信号的特征。简言之，为信号确定稳定性值/参数，并且基于所确定的稳定性值/参数来选择和应用适当的编码和/或解码模式。如本文所使用的，“编码模式”可以指编码模式和/或解码模式。如前所述，编码模式可以涉及用于处理信道错误和丢失分组的不同策略。此外，如本文所使用的，表达“解码模式”旨在指与音频信号的解码和重建相关联地使用的解码方法和/或用于错误隐藏的方法。也就是说，如本文所使用的，不同的解码模式可以与同一解码

方法、但与不同的错误隐藏方法相关联。类似地，不同的解码模式可以与同一错误隐藏方法、但不同的解码方法相关联。当应用于编解码器时，本文描述的解决方案涉及基于与音频信号稳定性相关的新颖测量来选择编码方法和/或错误隐藏方法。

[0052] 示例实施例

[0053] 下面，将参照图2a和图2b描述与用于对音频信号进行解码的方法相关的示例实施例。该方法可以由解码器执行，所述解码器可以被配置为符合音频解码的一个或多个标准。图2a中所示的方法包括：针对音频信号的帧m，在变换域中，确定201稳定性值D(m)。基于帧m的频谱包络的范围与相邻帧m-1的频谱包络的相应范围之间的差，确定该稳定性值D(m)。每个范围包括与音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。基于稳定性值D(m)，可以从多种解码模式中选择204解码模式。例如，可以选择解码方法和/或错误隐藏方法。然后可以应用205所选择的解码模式，来解码和/或重建音频信号中的至少帧m。

[0054] 如图所示，该方法还可以包括对稳定性值D(m)进行低通滤波202，从而获得滤波后的稳定性值$\tilde{D}(m)$。然后，可以通过使用例如S型函数将滤波后的稳定性值$\tilde{D}(m)$映射(203)至标量范围[0,1]，从而获得稳定性参数S(m)。然后，通过基于从D(m)导出的稳定性参数S(m)选择解码模式，实现基于D(m)的解码模式的选择。稳定性值的确定和稳定性参数的导出可以被认为是对音频信号的分段进行分类的方式，其中稳定性表示信号的某种类别或类型。

[0055] 作为示例，所描述的解码过程的适配可以涉及基于稳定性值从多个错误隐藏方法中选择错误隐藏方法。包括在例如解码器中的多个错误隐藏方法可以与单个解码方法相关联或与不同的解码方法相关联。如前所述，本文使用的术语解码模式可以指解码方法和/或错误隐藏方法。基于稳定性值或稳定性参数以及可能还基于其他标准，可以选择最适合于音频信号的相关部分的错误隐藏方法。稳定性值和参数可以指示音频信号的相关分段是否包括语音或音乐，和/或当音频信号包括音乐时：稳定性参数可以指示不同类型的音乐。错误隐藏方法中的至少一种可能更适合于语音而不是音乐，并且多个错误隐藏方法中的至少一种其他错误隐藏方法可能更适合于音乐而不是语音。然后，当稳定性值或稳定性参数(例如下面例示的那样，可能与进一步细化组合)指示音频信号的相关部分包括语音时，可以选择与音乐相比更适合于语音的错误隐藏方法。相应地，当稳定性值或参数指示音频信号的相关部分包括音乐时，可以选择与语音相比更适合于音乐的错误隐藏方法。

[0056] 这里描述的用于编解码器适配的方法的新颖性在于：使用音频信号的分段(在变换域中)的量化包络的范围来确定稳定性参数。相邻帧的包络的范围之间的差D(m)可以计算为：

[0057] $$D(m)=\sqrt{\frac{1}{b_{end}-b_{start}+1}\sum_{b=b_{start}}^{b_{end}}\left(E(m,b)-E(m-1,b)\right)^2}$$

[0058] 频带$b_{start},\cdots,b_{end}$表示用于包络差测量的频带的范围。它可以是连续的频带范围，或者，这些频带可以是不相交的，在这种情况下，表达b_{start}-b_{end}+1需要用该范围中的正确数量的频带来替代。注意，在针对第一帧的计算中，值E(m-1,b)不存在，并因此被初始化为例如与空频谱对应的包络值。

[0059] 执行对所确定的差D(m)的低通滤波以获得更稳定的控制参数。一种解决方案是：

使用以下形式的一阶AR(自回归)滤波器或遗忘因子：

[0060] $$\tilde{D}(m)=\alpha D(m)+(1-\alpha)D(m-1)$$

[0061] 其中α是AR滤波器的配置参数。

[0062] 为了促进在编解码器/解码器中使用滤波后的差或者稳定性值$\tilde{D}(m)$，可能希望将滤波后的差$\tilde{D}(m)$映射至更合适的使用范围。这里，如下使用S型函数将值$\tilde{D}(m)$映射至范围[0,1]：

[0063] $$S(m)=\frac{1}{1+e^{-b\left(\left(d-\tilde{D}(m)\right)+c\right)}}$$

[0064] 其中S(m)∈[0,1]表示映射的稳定性值。在示例性实施例中，可以将常数b,c,d设置为b＝6.11,c＝1.91和d＝2.26,但可以将b,c和d设置为任意合适的值。可以通过实验来设置S型函数的参数，使得其将观察到的输入参数$\tilde{D}(m)$的动态范围与期望的输出判定S(m)相适配。S型函数提供了用于实现软判决阈值的良好机制，原因在于可以控制拐点和工作范围二者。在图3a中示出了映射曲线，其中$\tilde{D}(m)$在水平轴上，S(m)在垂直轴上。由于指数函数在计算上是复杂的，因此可能期望用查找表替代映射函数。在那种情况下，如图3b中的圆圈所示，映射曲线将在$\tilde{D}(m)$和S(m)的成对的离散点中采样。在采样的情况下，如果优选的，$\tilde{D}(m)$和S(m)可以表示为例如$\hat{\tilde{D}}(m)$和$\hat{S}(m)$，在这种情况下，例如通过使用欧几里得距离定位$\tilde{D}(m)$的闭合值$\hat{\tilde{D}}(m)$来找到查找表值$\hat{S}(m)$。还可以注意到，由于函数的对称性，S型函数可以仅用过渡曲线的一半来表示。S型函数S_{mid}的中点定义为S_{mid}＝c/b+d。通过减去中点S_{mid}：

[0065] $$D'(m)=\left|\tilde{D}(m)-s_{mid}\right|$$

[0066] 我们可以如前所述使用量化和查找来获得相应的单侧映射稳定性参数，并且取决于相对于中点的位置导出的最终稳定性参数S′(m)为：

[0067] $$\begin{cases}\tilde{S}=1-D'(m), & \tilde{D}(m)<s_{mid}\\ \tilde{S}=D'(m), & \tilde{D}(m)\geq s_{mid}\end{cases}$$

[0068] 此外，可能期望对包络稳定性测量应用挂起(hangover)逻辑或滞后。也可能希望用瞬态检测器补充该测量。下面将进一步概述使用挂起逻辑的瞬态检测器的示例。

[0069] 另一实施例解决了产生包络稳定性测量的需要，其本身是更稳定的并且较少受统计波动的影响。如上所述，一种可能方案是对包络稳定性测量应用挂起逻辑或滞后。然而，在许多情况下，这可能不足够，并且另一方面，在一些情况下，仅产生具有有限数量的稳定性程度的离散输出就足矣。对于这种情况，已经发现使用采用马尔可夫模型的平滑器是有利的。这种平滑器将提供与通过对包络稳定性测量应用挂起逻辑或滞后可以实现的输出值相比更稳定(即更小的波动)的输出值。如果回头参考例如图2a和/或2b中的示例性实施例，

基于稳定性值或参数的解码模式选择(例如解码方法和/或错误隐藏方法)还可以基于定义与音频信号中的不同信号属性之间的转变相关的状态转移概率的马尔可夫模型。不同的状态可以例如表示语音和音乐。现在将描述使用马尔可夫模型产生具有有限数量的稳定性程度的离散输出的方法。

[0070] 马尔可夫模型

[0071] 所使用的马尔可夫模型包括M个状态,其中每个状态表示某一程度的包络稳定性。在M被选为2的情况下,一种状态(状态0)可以表示强烈波动的频谱包络,而另一种状态(状态1)可以表示稳定的频谱包络。在没有任何概念差异的情况下,有可能将这个模型扩展至更多状态,所述更多状态例如是针对中间包络稳定性程度的状态。

[0072] 该马尔可夫状态模型由状态转移概率表征,该状态转移概率表示从前一时刻的每个给定状态到当前时刻的给定状态的概率。例如,时刻可以与当前帧的帧索引m和先前正确接收的帧的帧索引m-1相对应。注意,在由于传输错误引起的帧丢失的情况下,这可能是与在没有帧丢失的情况下原本可用的先前帧不同的帧。状态转移概率可以以作为转移矩阵T的数学表达式来书写,其中每个元素表示当从状态i出现时转换至状态j的概率p(j|i)。对于优选的2状态马尔可夫模型,转移概率矩阵看起来如下。

[0073] $$T=\begin{bmatrix} p(0|0) & p(0|1) \\ p(1|0) & p(1|1) \end{bmatrix}$$。

[0074] 可以注意到,通过将保持在给定状态的似然性设置为相对较大的值,而将离开该状态的似然性设置较小的值,来实现期望的平滑效果。

[0075] 此外,每个状态与给定时刻的概率相关联。在先前正确接收帧m-1的时刻,状态概率由以下矢量给出:

[0076] $$P_S(m-1)=\begin{bmatrix} p_{S,0}(m-1) \\ p_{S,1}(m-1) \end{bmatrix}$$。

[0077] 为了计算每个状态的发生的先验似然性,将状态概率矢量$P_S(m-1)$乘以转移概率矩阵:

[0078] $P_A(m)=T\cdot P_S(m-1)$。

[0079] 然而,真实状态概率不仅取决于这些先验似然性,还取决于与当前帧时刻m处的当前观察$P_p(m)$相关联的似然性。根据本文介绍的实施例,要平滑的频谱包络测量值与这种观察似然性相关联。由于状态0表示波动频谱包络,并且状态1表示稳定包络,所以包络稳定性D(m)的低测量值意味着状态0的高概率和状态1的低概率。相反,如果测量或观察到的包络稳定性D(m)大,则这与状态1的高概率和状态0的低概率相关联。通过上述S型函数进行的包络稳定性测量值向良好适合于包络稳定性值的优选处理的状态观察似然性的映射是D(m)向状态1的状态观察概率的一一映射和1-D(m)向状态0的状态观察概率的一一映射。也就是说,S型函数映射的输出可以是马尔可夫平滑器的输入:

[0080] $$P_P(m)=\begin{bmatrix} p_{P,0}(m) \\ p_{P,1}(m) \end{bmatrix}=\begin{bmatrix} 1-D(m) \\ D(m) \end{bmatrix}$$。

[0081] 应当注意,该映射强烈依赖于所使用的S型函数。改变该功能可能需要将来自1-D

(m)和D(m)的映射函数引入各自的状态观察概率。除S型函数之外还可以进行的简单重映射是附加偏移和缩放因子的应用。

[0082] 在下一处理步骤中，状态观察概率矢量$P_P(m)$与先验概率矢量$P_A(m)$组合，先验概率矢量给出帧m的新状态概率矢量$P_S(m)$。这种组合通过两个矢量的逐元素相乘来完成：

[0083] $$\hat{P}_S(m)=\begin{bmatrix}\hat{p}_{S,0}(m)\\ \hat{p}_{S,1}(m)\end{bmatrix}=\begin{bmatrix}p_{P,0}(m)\cdot p_{A,0}(m)\\ p_{P,1}(m)\cdot p_{A,1}(m)\end{bmatrix}。$$

[0084] 由于该矢量的概率不一定总和为1，因此该矢量被重新归一化，这进而产生帧m的最终状态概率矢量：

[0085] $$P_S(m)=\frac{1}{\sum_i \hat{p}_{S,i}}\hat{P}_S(m)。$$

[0086] 在最后一个步骤中，通过该方法返回帧m的最可能状态作为平滑后的离散的包络稳定性测量。这需要标识状态概率矢量$P_S(m)$中的最大元素：

[0087] $$D_{smo}(m)=\max_i{}^{-1}(p_{S,i}(m))。$$

[0088] 为了使所描述的基于马尔可夫的平滑方法对于包络稳定性测量工作良好，以适当的方式选择状态转移概率。下面示出了已经发现非常适合于该任务的转移概率矩阵的示例：

[0089] $$T=\begin{bmatrix}0.999 & 0.5\\ 0.001 & 0.5\end{bmatrix}。$$

[0090] 从该转移概率矩阵中的概率可以看出，保持在状态0的可能性非常高，为0.999，而离开该状态的可能性很小，为0.001。因此，包络稳定性测量的平滑仅在包络稳定性测量值表示低稳定性的情况下是选择性的。由于指示稳定包络的稳定性测量值本身相对稳定，因此认为不需要对所述稳定性测量值进行进一步平滑。因此，离开状态1和停留在状态1中的转移似然值被等同地设置为0.5。

[0091] 要注意的是，通过增加状态的数量M可以容易地实现提高平滑后的包络稳定性测量的分辨率。

[0092] 包络稳定性测量的平滑方法的进一步增强的可能性是包括与包络稳定性呈现统计关系的附加测量。这种附加测量可以用模拟方式用作包络稳定性测量观察结果D(m)与状态观察概率的关联。在这种情况下，通过不同使用的测量的各自的状态观察概率的逐元素乘法来计算状态观察概率。

[0093] 已经发现，包络稳定性测量，特别是平滑后的测量，对于语音/音乐分类特别有用。根据该发现，语音可以与低稳定性测量良好地相关联，并且具体地与上述马尔可夫模型的状态0相关联。相反，音乐可以与高稳定性测量并且具体地与马尔可夫模型的状态1相关联。

[0094] 为了清楚起见，在具体实施例中，在每个时刻m，按下面的步骤执行上述平滑程序：

[0095] 1.将当前包络稳定性测量值D(m)与状态观察概率$P_P(m)$相关联。

[0096] 2.计算与较早时刻m-1的状态概率$P_S(m-1)$相关并与转移概率T相关的先验概率$P_A(m)$。

[0097] 3.将先验概率$P_A(m)$逐元素乘以状态观察概率$P_P(m)$，包括重新归一化，从而产生当前帧m的状态概率的矢量$P_S(m)$。

[0098] 4.标识状态概率矢量$P_S(m)$中具有最大概率的状态并且将其返回作为当前帧m的最终平滑后的包络稳定性测量$D_{smo}(m)$。

[0099] 图4是示出接收的音频帧的信号的频谱包络10的示意图形，其中每个频带的幅度由单个值表示。水平轴表示频率，并且垂直轴表示幅度(例如，功率等)。该图示出了针对较高频率增加带宽的典型设置，但是应当注意，可以使用任何类型的均匀或非均匀频带划分。

[0100] 瞬态检测

[0101] 如前所述，可能需要将稳定性值或稳定性参数与对音频信号的瞬态特征的测量结合。为了实现这种测量，可以使用瞬态检测器。例如，可以基于稳定性值/参数和瞬态测量来确定当对音频信号进行解码时应当使用哪种类型的噪声填充或衰减控制。下面概述了使用挂起逻辑的示例性瞬态检测器。术语“挂起”通常在音频信号处理中使用，并且指的是这样的想法：当通常认为延迟判决较安全时，延迟判决以避免在转换时段中的不稳定切换行为。

[0102] 瞬态检测器根据编码模式使用不同的分析。它具有挂起计数器no_att_hangover以处理挂起逻辑，该挂起逻辑被初始化为零。瞬态检测器具有针对三种不同模式定义的行为：

[0103] 模式A不具有包络值的低频带编码模式

[0104] 模式B具有包络值的正常编码模式

[0105] 模式C瞬态编码模式

[0106] 瞬态检测依赖于合成信号的长期能量估计。它根据编码模式进行不同更新。

[0107] 模式A

[0108] 在模式A中，帧能量估计$E_{frameA}(m)$被计算为

[0109] $$E_{frameA}(m)=\sqrt{\frac{1}{bin_th}\sum_{k=0}^{bin_th}\hat{X}(m,k)^2}$$

[0110] 其中bin_th是模式A的合成低频带中的最高编码系数，并且$\hat{X}(m,k)$是帧m的合成的MDCT系数。在编码器中，这些是使用可以在编码处理中提取的本地合成方法来再现的，并且它们与在解码处理中获得的系数相同。使用低通滤波器更新长期能量估计E_{LT}

[0111] $E_{LT}(m)=\beta E_{LT}(m-1)+(1-\beta)E_{frameA}(m)$

[0112] 其中β是滤波因子，该滤波因子具有示例性值0.93。如果挂起计数器大于1，则它被递减。

[0113] $$\begin{cases} no_att_hangover(m)=no_att_hangover(m-1)-1, & no_att_hangover>0 \\ no_att_hangover(m)=no_att_hangover(m-1), & no_att_hangover=0 \end{cases}$$

[0114] 模式B

[0115] 基于量化的包络值来更新长期能量估计$E_{frameB}(m)$。

[0116] $$E_{frameB}(m)=\sum_{b=0}^{B_{LF}}\hat{E}(m,b)$$

[0117] 其中B_{LF}是在低频能量计算中包括的最高频带b。以与模式A中相同的方式更新该长期能量估计：

[0118] $E_{LT}(m)=\beta E_{LT}(m-1)+(1-\beta)E_{frameB}(m)$

[0119] 与模式A等同地执行挂起递减。

[0120] 模式C

[0121] 模式C是在四个子帧(每个子帧对应于LTE中的1ms)中对频谱进行编码的瞬态模式。包络被交织成这样的样式,该样式中频率阶的一部分被保持。根据以下计算四个子帧能量$E_{sub,SF}$,SF=0,1,2,3:

[0122] $$E_{sub,SF}(m)=\frac{1}{|subframeSF|}\sum_{b\in subframeSF}\hat{E}(m,b)$$

[0123] 其中subframeSF表示代表子帧SF的包络频带b,并且|subframeSF|是该集合的大小。注意,实际实施方式将取决于交织后的子帧在包络矢量中的布置。

[0124] 帧能量$E_{frameC}(m)$通过对子帧能量求和来形成:

[0125] $$E_{frameC}(m)=\sum_{sf=0}^{3}E_{sub,sf}(m)$$

[0126] 通过检查下述条件,对高能量帧进行瞬态测试

[0127] $E_{frameC}(m)>E_{THR}\cdot N_{SF}$

[0128] 其中E_{THR}=100是能量阈值,并且N_{SF}=4是子帧的数量。如果通过了上述条件,则找到最大子帧能量差:

[0129] $$D_{\max}(m)=\max_{SF}\frac{\left(E_{sub,SF}(m)-E_{sub,SF-1}(m)\right)}{E_{LT}(m)},SF=0,1,2,3$$

[0130] 最后,如果条件$D_{max}(m)>D_{THR}$为真(其中D_{THR}=5是取决于实施方式和灵敏度设置的判决阈值),则将挂起计数器设置为最大值

[0131] $$\begin{cases}no_att_hangover(m)=no_att_hangover(m-1)-1,\\ no_att_hangover>0\wedge D_{\max}(m)\le D_{THR}\\ no_att_hangover(m)=ATT_LIM_HANGOVER,\\ D_{\max}(m)>D_{THR}\end{cases}$$

[0132] 其中ATT_LIM_HANGOVER=150是可配置的恒定帧计数器值。现在,如果条件T(m)=no_att_hangover(m)>0为真,则意味着已经检测到瞬态,并且挂起计数器尚未达到零。

[0133] 瞬时挂起判决T(m)可以与包络稳定性测量$\tilde{S}(m)$组合,使得仅在T(m)为真时,应用该取决于$\tilde{S}(m)$的修改。

[0134] 一个特定问题是在不提供具有子带范数(或缩放因子)形式的频谱包络表示的音频编解码器的情况下包络稳定性测量的计算。

[0135] 下面描述一个实施例,其解决了该问题,并且仍然获得与如上所述的基于子带范数或缩放因子获得的包络稳定性测量一致的有用的包络稳定性测量。

[0136] 该解决方案的第一步是找到给定信号帧的频谱包络的合适的替代表示。一种这种表示是基于线性预测系数(LPC或短期预测系数)的表示。如果适当地选择LPC阶数P,例如对于宽带或超宽带信号为16,则这些系数是频谱包络的良好表示。特别适合于编码、量化和内插目的的LPC参数的表示是线谱频率(LSF)或相关参数,例如ISF(导抗谱频率)或LSP(线谱

对)。原因在于这些参数表现出与相应的LPC合成滤波器的包络谱的良好的关系。

[0137] 评估当前帧的LSF参数与先前帧的LSF参数相比的稳定性的现有技术度量在ITU-T G.718编解码器中被称为LSF稳定性度量。该LSF稳定性度量用于LPC参数内插的上下文中以及帧擦除的情况下。该度量定义如下：

[0138] $$lsf_stab(m) = a - b \cdot \sum_{i=1}^{P} \left(lsf_i(m) - lsf_i(m-1)\right)^2 ,$$

[0139] 其中P是LPC滤波器阶数,a和b是某些合适的常数。此外,lsf_stab度量可以限制为从0至1的间隔。接近1的大数意味着LSF参数非常稳定,即没有太大变化,而低值意味着参数相对不稳定。

[0140] 根据本文介绍的实施例的一个发现是,LSF稳定性度量还可以用作包络稳定性的特别有用的指示符,作为以子带范数(或缩放因子)的形式比较当前和较早的频谱包络的备选。为此,根据一个实施例,针对当前帧(相对于较早的帧)计算lsf_stab参数。然后,通过适当的多项式变换来重新缩放该参数,如：

[0141] $$\hat{D}(m) = \sum_{n=0}^{N} \alpha_n \left(lsf_stab(m)\right)^n ,$$

[0142] 其中N是多项式阶数,并且α_n是多项式系数。

[0143] 执行重新缩放,即多项式阶数和系数的设置,使得变换值$\hat{D}(m)$尽可能与上述的相应包络稳定性值D(m)相类似地工作。发现在许多情况下多项式阶数为1足矣。

[0144] 分类,图5a和5b

[0145] 上述方法可以被描述为用于对音频信号的一部分进行分类的方法,并且其中可以基于分类的结果选择适当的解码或编码模式或方法。

[0146] 图5a-b是示出在主机设备(例如,图1的无线终端和/或转码节点)的音频编码器中执行的、用于辅助针对音频的编码模式的选择的方法的流程图。

[0147] 在获得编解码器参数步骤501中,可以获得编解码器参数。编解码器参数是主机设备的编码器或解码器中已经可用的参数。

[0148] 在分类步骤502中,基于编解码器参数对所述音频信号分类。可以例如分类为语音或音乐。可选地,如以上更详细地解释的,在该步骤中使用滞后以防止来回跳跃。附加地或备选地,如上文更详细解释的,马尔可夫模型(例如马尔可夫链)可用于提高分类的稳定性。

[0149] 例如,分类可以基于音频数据的频谱信息的包络稳定性测量,然后在该步骤中计算。该计算可以例如基于量化包络值。

[0150] 可选地,该步骤包括将稳定性测量映射至预定义的如上面的S(m)表示的标量范围,可选地使用查找表来降低计算要求。

[0151] 可以针对音频数据的每个接收帧重复该方法。

[0152] 图5b示出了根据一个实施例的用于辅助对音频的编码和/或解码模式的选择的方法。该方法与图5a中所示的方法类似,并且将仅描述相对于图5a而言的新的或修改的步骤。

[0153] 在可选的选择编码模式步骤503中,基于来自分类步骤502的分类来选择编码模式。

[0154] 在可选的编码步骤504中,基于在选择编码模式步骤503中选择的编码模式对音频

数据进行编码或解码。

[0155] 具体实现

[0156] 上述方法和技术可以在编码器和/或解码器中实现，编码器和/或解码器可以是例如通信设备的一部分。

[0157] 解码器，图6a-6c

[0158] 在图6a中以一般方式示出了解码器的示例实施例。解码器指的是：被配置用于解码和可能以其他方式重建音频信号的解码器。解码器有可以还被配置用于对其他类型的信号进行解码。解码器600被配置为执行上述例如参照图2a和图2b的方法实施例中的至少一个。解码器600关联于与前述方法实施例相同的技术特征、目的和优点。解码器可以被配置为符合音频编码/解码的一个或多个标准。为了避免不必要的重复，将简要描述解码器。

[0159] 可以如下实现和/或描述解码器：

[0160] 解码器600被配置为对音频信号进行解码。解码器600包括处理电路或处理装置601和通信接口602。处理电路601被配置为：针对帧m，在变换域中，使所述解码器600：基于帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差，确定稳定性值D(m)，每个范围包括与所述音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。处理电路601还被配置为：使所述解码器基于所述稳定性值D(m)从多个解码模式中选择编码模式；以及应用所选的解码模式。

[0161] 处理电路601还可以被配置为使解码器对稳定性值D(m)进行低通滤波，从而获得滤波后的稳定性值$\widetilde{D}(m)$；以及通过使用S型函数将滤波后的稳定性值$\widetilde{D}(m)$映射至标量范围[0,1]，从而获得稳定性参数S(m)，然后基于稳定性参数S(m)选择解码模式。通信接口602，其也可以被标记为例如输入/输出(I/O)接口，其包括用于向其它实体或模块发送数据和从其他实体或模块接收数据的接口。

[0162] 如图6b中所示，处理电路601可以包括处理装置(例如处理器603(例如CPU))和用于存储或保持指令的存储器604。然后，存储器将包括例如具有计算机程序605的形式的指令，所述指令当由处理装置603执行时，使解码器600执行上述动作。

[0163] 在图6c中示出了处理电路601的备选实现方式。这里处理电路包括编码单元606，被配置为使所述解码器600确定以下关系：基于帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差，确定稳定性值D(m)，每个范围包括与所述音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。处理电路还包括选择单元609，被配置为使所述解码器基于稳定性值D(m)从多个解码模式中选择解码模式。处理电路还包括应用单元或解码单元610，被配置为使所述解码器应用所选的解码模式。处理电路601可以包括更多单元，例如滤波单元607，被配置为使解码器对稳定性值D(m)进行低通滤波，从而获得滤波后的稳定性值$\widetilde{D}(m)$处理电路还可以包括映射单元608，其被配置为使所述解码器通过使用S型函数将滤波后的稳定性值$\widetilde{D}(m)$映射至标量范围[0,1]，从而获得稳定性参数S(m)，然后基于稳定性参数S(m)选择解码模式。在图6c中用虚线轮廓示出这些可选单元。

[0164] 上述解码器或编解码器可以被配置用于本文所述的不同方法实施例，例如使用马尔可夫模型并在与错误隐藏相关联的不同的解码模式之间进行选择的方法实施例。

[0165] 可以假设编码器600包括用于执行常规解码器功能的附加功能。

[0166] 编码器,图7a-7c

[0167] 在图7a中以一般方式示出了编码器的示例实施例。编码器指被配置用于对音频信号进行编码的编码器。编码器可以还被配置为对其他类型的信号进行编码。编码器700被配置为执行与上述例如参照图2a和图2b所描述的解码方法相对应的至少一种方法。也就是说,替代于选择解码模式(如图2a和图2b所示),选择并应用编码模式。编码器700关联于与前述方法实施例相同的技术特征、目的和优点。编码器可以被配置为符合音频编码/解码的一个或多个标准。为了避免不必要的重复,将简要描述编码器。

[0168] 可以如下实现和/或描述编码器:

[0169] 编码器700被配置为对音频信号进行编码。编码器700包括处理电路或处理装置701和通信接口702。处理电路701被配置为:在变换域中,针对帧m,使所述编码器700:基于帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差,确定稳定性值D(m),每个范围包括与所述音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。处理电路701还被配置为使所述编码器基于稳定性值D(m)从多个编码模式中选择编码模式,以及应用所选的编码模式。

[0170] 处理电路701还可以被配置为使编码器对稳定性值D(m)进行低通滤波,从而获得滤波后的稳定性值$\widetilde{D}(m)$;以及通过使用S型函数将滤波后的稳定性值$\widetilde{D}(m)$映射至标量范围[0,1],从而获得稳定性参数S(m),然后基于稳定性参数S(m)选择编码模式。通信接口702,其也可以被标记为例如输入/输出(I/O)接口,其包括用于向其它实体或模块发送数据和从其他实体或模块接收数据的接口。

[0171] 如图7b所示,处理电路701可以包括处理装置(例如处理器703(例如CPU))和用于存储或保持指令的存储器704。然后,存储器将包括例如具有计算机程序705的形式的指令,所述指令当由处理装置703执行时,使编码器700执行上述动作。

[0172] 在图7c示出了处理电路701的备选实施方式。这里处理电路包括编码单元706,被配置为使所述编码器700确定以下关系:基于帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差确定稳定性值D(m),每个范围包括与所述音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。处理电路还包括选择单元709,被配置为使所述编码器基于稳定性值D(m)从多个编码模式中选择编码模式。处理电路还包括应用单元或编码单元710,被配置为使所述编码器应用所选的编码模式。处理电路701可以包括更多单元,例如滤波单元707,被配置为使编码器对稳定性值D(m)进行低通滤波,从而获得滤波后的稳定性值$\widetilde{D}(m)$。处理电路还可以包括映射单元708,被配置为使所述编码器通过使用S型函数将滤波后的稳定性值$\widetilde{D}(m)$映射至标量范围[0,1],从而获得稳定性参数S(m),然后基于稳定性参数S(m)选择解码模式。这些可选单元在图7c中用虚线轮廓示出。

[0173] 上述编码器或编解码器可以被配置用于本文所述的不同方法实施例,例如使用马尔可夫模型的方法实施了。

[0174] 可以认为编码器700包括用于执行常规编码器功能的附加功能。

[0175] 分类器,图8a-8c

[0176] 在图8a中以一般方式示出了分类器的示例实施例。分类器是指被配置用于对音频信号进行分类，即区分不同类型或类别的音频信号的分类器。分类器800被配置为执行与上述例如参照图5a和图5b所描述的方法相对应的至少一种方法。分类器800关联于与前述方法实施例相同的技术特征、目的和优点。分类器可以被配置为符合音频编码/解码的一个或多个标准。为了避免不必要的重复，将简要描述分类器。

[0177] 可以如下实现和/或描述分类器：

[0178] 分类器800被配置为对音频信号进行分类。分类器800包括处理电路或处理装置801和通信接口802。处理电路801被配置为：在变换域中，针对帧m，使所述分类器800：基于帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差，确定稳定性值D(m)，每个范围包括与所述音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。处理电路801还被配置为使分类器基于稳定性值D(m)对音频信号进行分类。例如，分类可以涉及从多个候选音频信号类中选择音频信号类。处理电路801还可以被配置为使分类器指示例如由解码器或编码器使用的分类。

[0179] 处理电路801还可以被配置为使分类器对稳定性值D(m)进行低通滤波，从而获得滤波后的稳定性值$\widetilde{D}(m)$；以及通过使用S型函数将滤波后的稳定性值$\widetilde{D}(m)$映射至标量范围[0,1]，从而获得稳定性参数S(m)，可以基于稳定性参数S(m)对音频信号进行分类。通信接口802，其也可以被标记为例如输入/输出(I/O)接口，其包括用于向其它实体或模块发送数据和从其他实体或模块接收数据的接口。

[0180] 如图8b中所示，处理电路801可以包括处理装置(例如处理器803(例如CPU))和用于存储或保持指令的存储器804。然后，存储器将包括例如具有计算机程序805形式的指令，所述指令当由处理装置803执行时，使分类器800执行上述动作。

[0181] 在图8c中示出了处理电路801的备选实施方式。这里处理电路包括编码单元806，被配置为使所述分类器800确定以下关系：基于帧m的频谱包络的范围和相邻帧m-1的频谱包络的相应范围之间的差，确定(201)稳定性值D(m)，每个范围包括与所述音频信号的分段的频谱带中的能量相关的量化频谱包络值的集合。处理电路还包括分类单元809，被配置为使所述分类器对所述音频信号分类。处理电路还可以包括指示单元810，被配置为使得分类器向例如编码器或解码器指示分类。处理电路801可以包括更多单元，例如滤波单元807，被配置为使分类器对稳定性值D(m)进行低通滤波，从而获得滤波后的稳定性值$\widetilde{D}(m)$。处理电路还可以包括映射单元808，被配置为使所述分类器通过使用S型函数将滤波后的稳定性值$\widetilde{D}(m)$映射至标量范围[0,1]，从而获得稳定性参数S(m)，然后可以基于稳定性参数S(m)对音频信号进行分类。在图8c中用虚线轮廓示出这些可选单元。

[0182] 上述分类器可以被配置用于本文所述的不同方法实施例，例如使用马尔可夫模型的方法实施例。

[0183] 可以假设分类器800包括用于执行常规分类器功能的附加功能。

[0184] 图9是示出了图1的无线终端2的一些组件的示意图。使用合适的中央处理单元(CPU)、多处理器、微控制器、数字信号处理器(DSP)、专用集成电路等等中的一个或多个的任意组合提供处理器70，处理器70能够执行在存储器74中存储的软件指令76，软件指令76

因此可以是计算机程序产品。处理器70可以执行软件指令76,以执行以上参照图5a-b所描述的方法的一个或多个实施例。

[0185] 存储器74可以是读写存储器(RAM)和只读存储器(ROM)的任意组合。存储器74还包括持久存储设备,其例如可以是磁存储器、光存储器、固态存储器或甚至远程安装存储器中的任意单独一个或其组合。

[0186] 还提供数据存储器73,用于在处理器70中执行软件指令期间读取和/或存储数据。数据存储器73可以是读写存储器(RAM)和只读存储器(ROM)的任意组合。

[0187] 无线终端2还包括I/O接口72,用于与其他外部实体通信。I/O接口72还包括用户接口,包括麦克风、扬声器、显示器等。可选地,外部麦克风和/或扬声器/头戴耳机可以与无线终端连接。

[0188] 无线终端2还包括一个或多个收发机71,包括模拟和数字组件和合适数量的天线75,用于与图1中所示的无线终端进行无线通信。

[0189] 无线终端2包括音频编码器和音频解码器。这些可以用软件指令76实现,软件指令76可以由处理器70执行或使用单独硬件执行(未示出)。

[0190] 为了突出本文介绍的构思,省略了无线终端2的其他组件。

[0191] 图10是示出了图1的转码节点5的一些组件的示意图。使用合适的中央处理单元(CPU)、多处理器、微控制器、数字信号处理器(DSP)、专用集成电路等等中的一个或多个的任意组合提供处理器80,处理器80能够执行在存储器84中存储的软件指令66,软件指令66因此可以是计算机程序产品。处理器80可以被配置为执行软件指令86,以执行以上参照图5a-b所描述的方法的一个或多个实施例。

[0192] 存储器84可以是读写存储器(RAM)和只读存储器(ROM)的任意组合。存储器84还包括持久存储设备,其例如可以是磁存储器、光存储器、固态存储器或甚至远程安装存储器中的任意单独一个或其组合。

[0193] 还提供数据存储器83,用于在处理器80中执行软件指令期间读取和/或存储数据。数据存储器83可以是读写存储器(RAM)和只读存储器(ROM)的任意组合。

[0194] 转码节点5还包括I/O接口82,用于经由无线电基站1与其他外部实体(例如图1的无线终端)通信。

[0195] 转码节点5包括音频编码器和音频解码器。这些可以用软件指令86实现,软件指令86可以由处理器80执行或使用单独硬件执行(未示出)。

[0196] 为了突出本文介绍的构思,省略了转码节点5的其他组件。

[0197] 图11示出了包括计算机可读装置的计算机程序产品90的一个示例。在该计算机可读装置中,可以存储计算机程序91,该计算机程序可以使处理器执行根据本文描述的实施例的方法。在本示例中,计算机程序产品是光盘,例如CD(紧致盘)或DVD(数字多功能盘)或蓝光盘。如以上解释的,计算机程序产品还可以被实现为设备的存储器,例如图7的计算机程序产品74或图8的计算机程序产品84。虽然计算机程序91此处被示意性地示为光盘上所示的轨道,可以用任意适合于计算机程序产品的方式来存储计算机程序,例如可移除固态存储器(例如,通用串行总线(USB)棒)。

[0198] 这里,现在以下是一组编号的实施例,以进一步例示本文所介绍的发明构思的一些方面。

[0199] 1.一种用于辅助对音频的编码模式或解码模式的选择的方法,所述方法在音频编码器或解码器中执行,并且包括步骤:

[0200] 获得(501)编解码器参数;以及

[0201] 基于所述编解码器参数对所述音频信号分类(502)。

[0202] 2.根据实施例1所述的方法,还包括步骤:

[0203] 基于所述分类来选择(503)编码模式。

[0204] 3.根据实施例2所述的方法,还包括以下步骤:

[0205] 基于在选择步骤中选择的编码模式对音频数据进行编码或解码(504)。

[0206] 4.根据前述实施例中任一实施例所述的方法,其中对所述音频信号进行分类(502)的步骤包括使用滞后。

[0207] 5.根据前述实施例中任一实施例所述的方法,其中对所述音频信号进行分类(502)的步骤包括使用马尔可夫链。

[0208] 6.根据前述实施例中任一实施例所述的方法,其中所述分类步骤(502)包括计算音频数据的频谱信息的包络稳定性测量。

[0209] 7.根据实施例6所述的方法,其中,在分类步骤中,包络稳定性测量的计算基于量化包络值。

[0210] 8.根据实施例6或实施例7所述的方法,其中所述分类步骤包括将所述稳定性测量映射至预定义的标量范围。

[0211] 9.根据实施例8所述的方法,其中,所述分类步骤包括使用查找表将所述稳定性测量映射至预定义的标量范围。

[0212] 10.根据前述实施例中任一实施例所述的方法,其中所述包络稳定性测量基于帧m中的包络特征与先前帧m-1中的包络特征的比较。

[0213] 11.一种用于辅助对音频编码模式的选择的主机设备(2,5),所述主机设备包括:

[0214] 处理器(70,80);以及

[0215] 存储器(74,84),用于存储指令(76,86),所述指令当由处理器(21)执行时,使所述主机设备(2,5):

[0216] 获得编解码器参数;以及

[0217] 基于所述编解码器参数对所述音频信号分类。

[0218] 12.根据实施例11所述的主机设备(2,5),还包括以下指令,所述指令当由所述处理器执行时,使所述主机设备(2,5)基于所述分类来选择编码模式。

[0219] 13.根据实施例12所述的主机设备(2,5),还包括以下指令,所述指令当由所述处理器执行时,使所述主机设备(2,5)基于所选的编码模式对语音数据进行编码。

[0220] 14.根据实施例11至13中任一项所述的主机设备(2,5),其中用于对音频信号进行分类的指令还包括以下指令,所述指令当由所述处理器执行时,使所述主机设备(2,5)使用滞后。

[0221] 15.根据实施例11至14中任一项所述的主机设备(2,5),其中用于对音频信号进行分类的指令包括以下指令,所述指令当由所述处理器执行时,使所述主机设备(2,5)使用马尔可夫链。

[0222] 16.根据实施例11至15中任一项所述的主机设备(2,5),其中用于分类的指令包括

以下指令,所述指令当由所述处理器执行时,使所述主机设备(2,5)计算语音数据的频谱包络的包络稳定性测量。

[0223] 17.根据实施例16所述的主机设备(2,5),其中用于分类的指令包括以下指令,所述指令当由所述处理器执行时,使所述主机设备(2,5)基于量化包络值来计算包络稳定性测量。

[0224] 18.根据实施例16或实施例17所述的主机设备(2,5),其中用于分类的指令包括以下指令,所述指令当由所述处理器执行时,使所述主机设备(2,5)将所述稳定性测量映射至预定标量范围。

[0225] 19.根据实施例18所述的主机设备(2,5),其中用于分类的指令包括以下指令,所述指令当由所述处理器执行时,使所述主机设备(2,5)使用查找表将稳定性测量映射值预定标量范围。

[0226] 20.根据实施例11至19中任一项所述的主机设备(2,5),其中用于分类的指令包括以下指令,所述指令当由所述处理器执行时,使所述主机设备(2,5)基于帧m中的包络特征与前一帧m-1中的包络特征的比较来计算包络稳定性测量。

[0227] 21.一种计算机程序(66,91),用于辅助对音频的编码模式的选择,所述计算机程序包括计算机程序代码,所述计算机程序代码在主机设备上运行时使所述主机设备(2,5):

[0228] 获得编解码器参数;以及

[0229] 基于所述编解码器参数对所述音频信号分类。

[0230] 22.一种计算机程序产品(74,84,90),包括:根据实施例21所述的计算机程序以及其上存储所述计算机程序的计算机可读装置。

[0231] 已经参考一些实施例在上文中主要地描述了本发明。然而,本领域技术人员容易理解,在本发明的范围中的不同于上文公开的实施例的其他实施例也是可能的。

[0232] 结论陈述

[0233] 本文所述的步骤、功能、过程、模块、单元和/或框可以使用任何常规技术实现在硬件中,例如使用分立式电路或集成电路技术,包括通用电子电路和专用电路二者。

[0234] 特定示例包括一个或多个合适配置的数字信号处理器和其他已知电子电路,例如用于执行特定功能的互连的分立逻辑门、或者专用集成电路(ASIC)。

[0235] 备选地,上述步骤、功能、过程、模块、单元和/或框的至少一些可以实现在软件中,所述软件例如是由包括一个或多个处理单元的合适的处理电路来执行的计算机程序。在网络节点中的计算机程序的使用之前和/或期间,软件可以由诸如电子信号,光学信号,无线电信号或计算机可读存储介质的载体承载。上述网络节点和索引服务器可以在所谓的云解决方案中实现,指的是该实施方式可以是分布式的,并且因此网络节点和索引服务器可以是所谓的虚拟节点或虚拟机。

[0236] 当由一个或多个处理器执行时,本文介绍的流程图(一个或多个)可被认为是计算机流程图(一个或多个)。对应的装置可定义为一组功能模块,其中由处理器执行的每个步骤对应于一个功能模块。在这种情况下,功能模块实现为在处理器上运行的计算机程序。

[0237] 处理电路的示例包括但不限于:一个或多个微处理器、一个或多个数字信号处理器(DSP)、一个或多个中央处理单元(CPU)和/或任意合适的可编程逻辑电路,例如一个或多个现场可编程门阵列(FPGA)或者一个或多个可编程逻辑控制器(PLC)。也就是说,在上述不

同节点中的布置中的单元或模块可以被实施为模拟或数字电路的组合、和/或由存储在存储器中的软件和/或固件配置的一个或多个处理器。这些处理器中的一个或多个以及其它数字硬件可以包括在单个专用集成电路(ASIC)中,或者若干个处理器和各种数字硬件可以分布在若干个分离的组件上,不论是单独封装的还是组装为片上系统(SoC)的。

[0238] 也应当理解,可以重用实现了所提出的技术的任意传统设备或单元的通用处理能力。还可以例如通过重新编程现有的软件或者通过添加新的软件组件来重新使用现有的软件。

[0239] 仅作为示例,提出上述实施例,并且应当理解,所提出的技术不限于此。本领域技术人员将会理解,在不背离本发明范围的情况下,可以对该实施例进行各种修改、合并和改变。尤其是,在其他技术上可行的配置中,不同实施例中的不同部分的方案可以进行组合。

[0240] 当使用词语"包括"或"包括……"时,应当被理解为非限制性的,即意味着"至少包括"。

[0241] 应该注意的是,在一些替代实施例中,在框中标记的功能/动作可以不以流程图中标记的顺序发生。例如依赖于所涉及的功能/动作,连续示出的两个框实际上可以实质上同时执行,或者框有时候可以按照相反的顺序执行。此外,可以将流程图和/或方框图中的给定模块的功能分离成多个框和/或流程图的两个或更多框的功能和/或可以至少部分地集成方框图。最后,在不脱离本发明构思的范围的情况下,可以在所示出的块之间添加/插入其他块,和/或可以省略块/操作。

[0242] 应当理解,在本公开内对交互单元的选择以及单元的命名仅用于示例目的,并且可通过多个备选方式来配置适合于执行上述任何方法的节点,从而能够执行所公开的处理动作。

[0243] 还应当注意,本公开中描述的单元应被认为是逻辑实体,而不必是分离的物理实体。

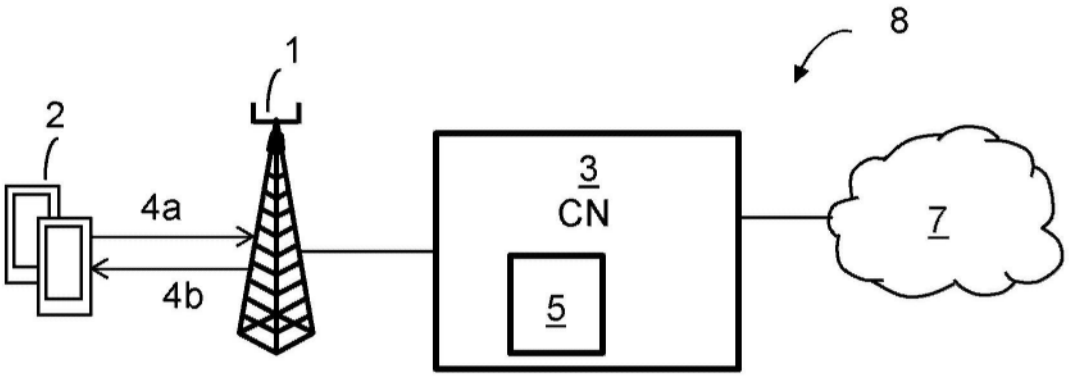

8
1
2
4a
4b
3
CN
5
7

图1

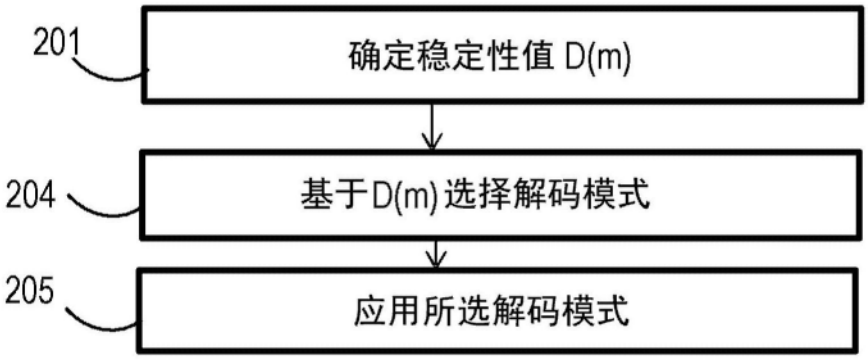

201
确定稳定性值 D(m)
204
基于D(m)选择解码模式
205
应用所选解码模式

图2a

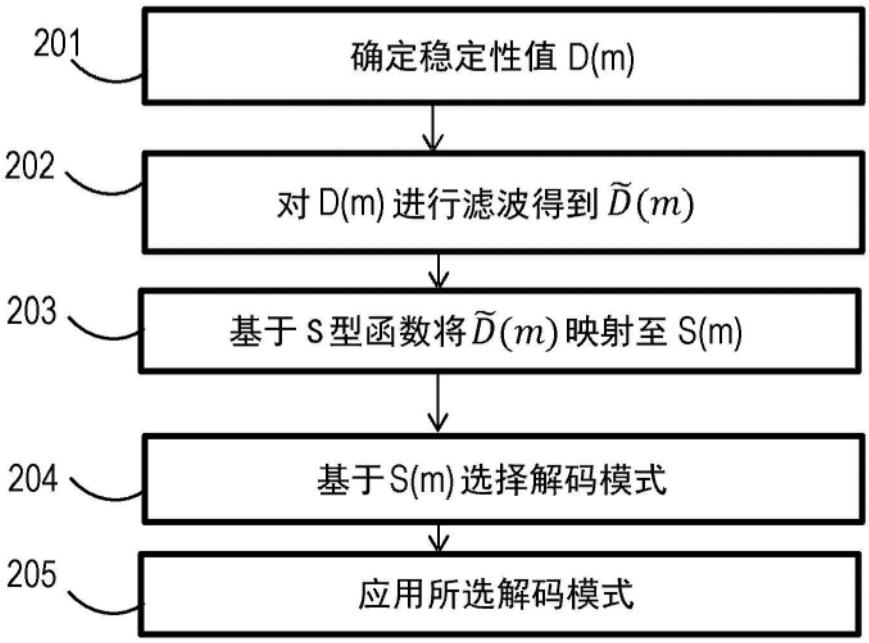

201
确定稳定性值 D(m)
202
对 D(m) 进行滤波得到 $\widetilde{D}(m)$
203
基于 S 型函数将 $\widetilde{D}(m)$ 映射至 S(m)
204
基于S(m)选择解码模式
205
应用所选解码模式

图2b

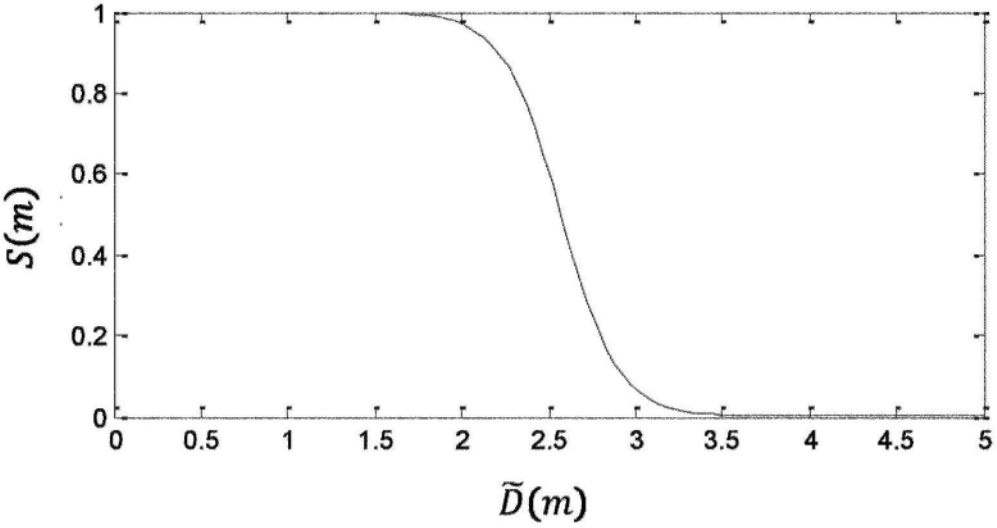

1
0.8
0.6
0.4
0.2
0
S(m)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
$\tilde{D}(m)$

图3a

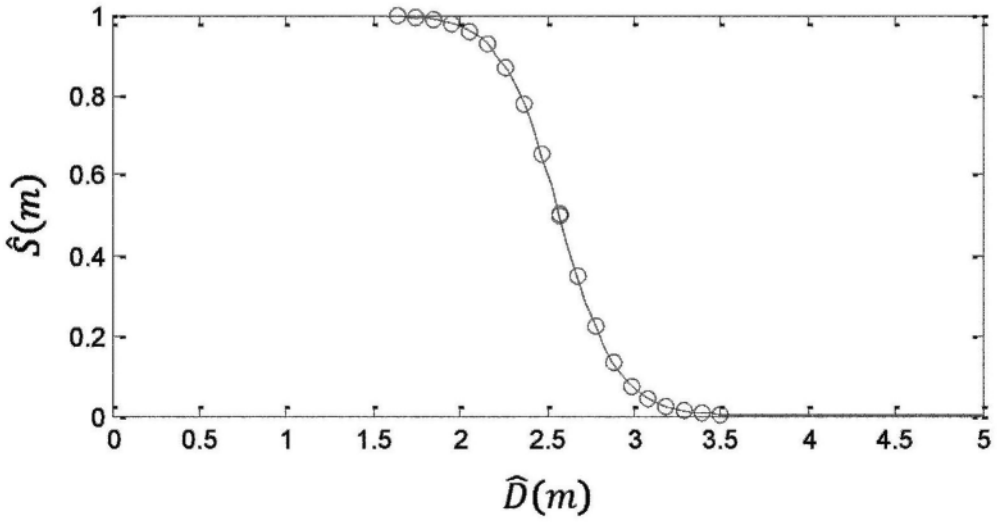

1
0.8
0.6
0.4
0.2
0
$\hat{S}(m)$
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
$\hat{D}(m)$

图3b

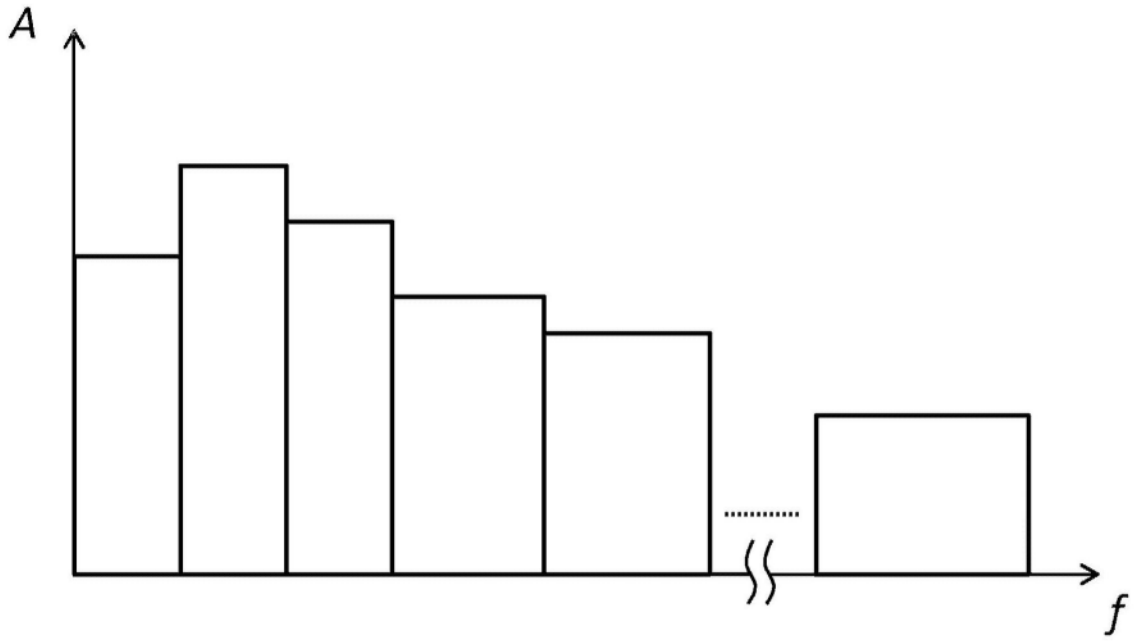

A
f

图4

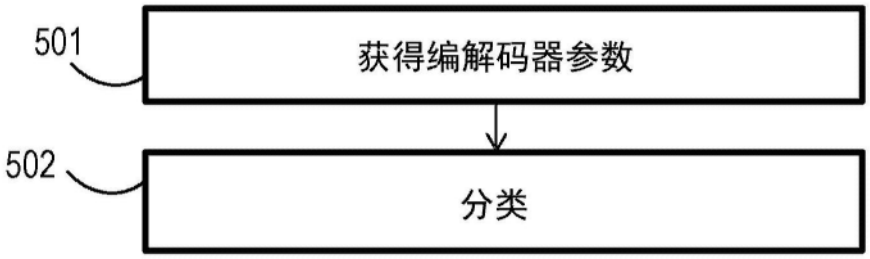

501
获得编解码器参数
502
分类

图5a

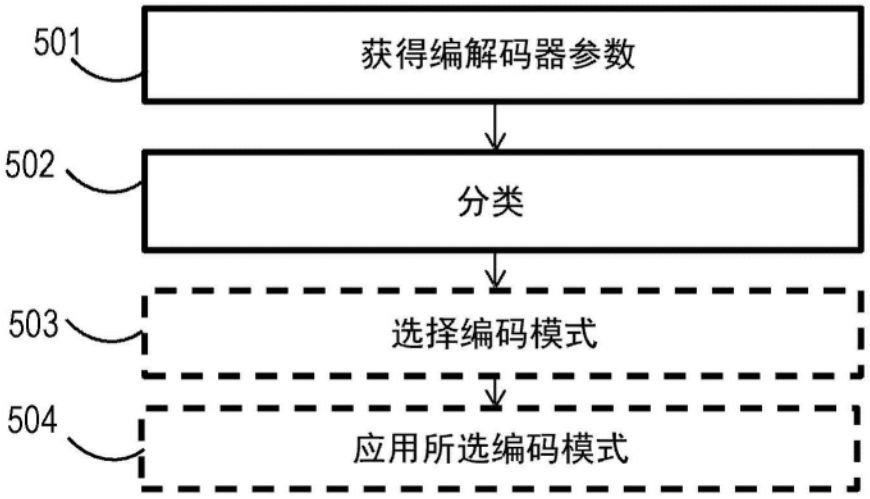

501
获得编解码器参数
502
分类
503
选择编码模式
504
应用所选编码模式

图5b

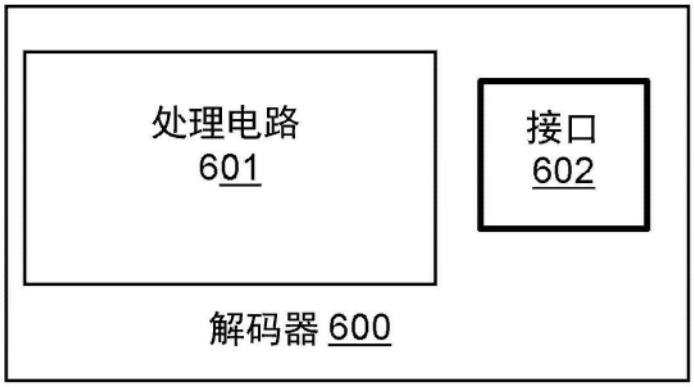

处理电路
601
接口
602
解码器 600

图6a

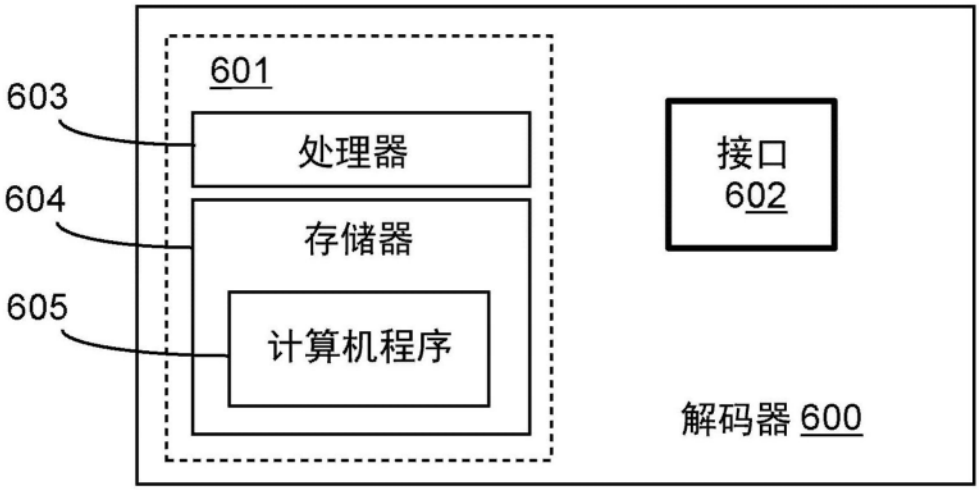

601
603
处理器
604
存储器
605
计算机程序
接口
602
解码器 600

图6b

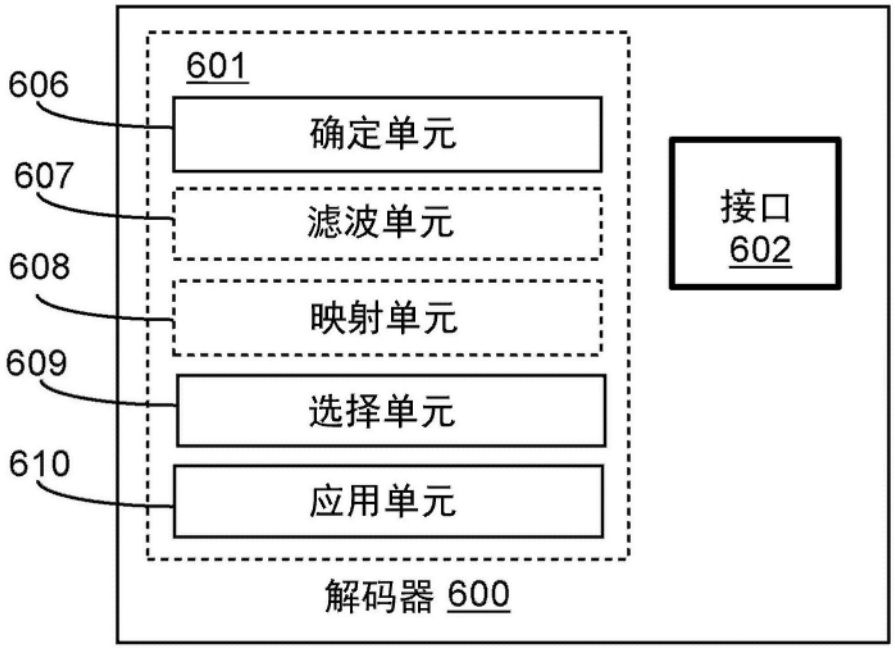

601
606
确定单元
607
滤波单元
608
映射单元
609
选择单元
610
应用单元
接口
602
解码器 600

图6c

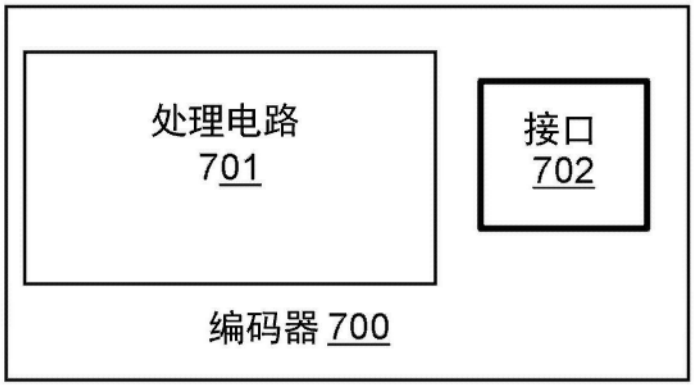

处理电路
701
接口
702
编码器 700

图7a

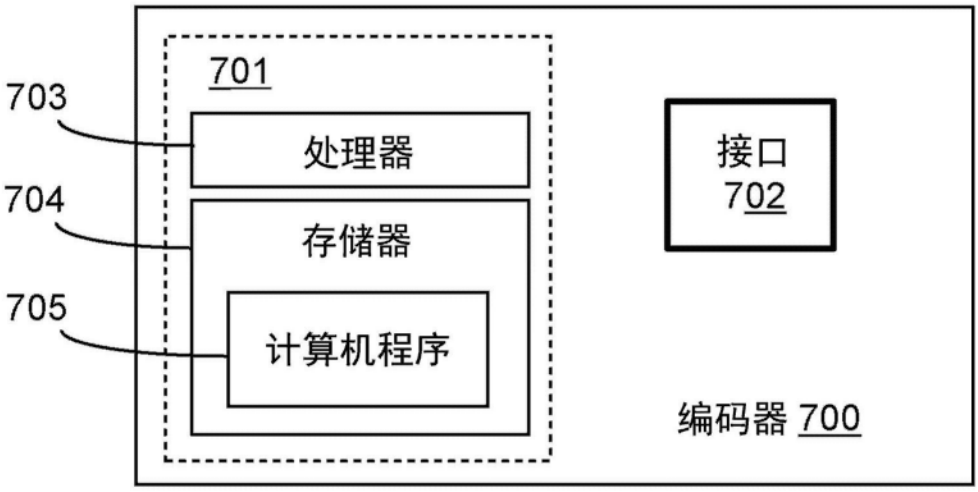

701
703
处理器
704
存储器
705
计算机程序
接口
702
编码器 700

图7b

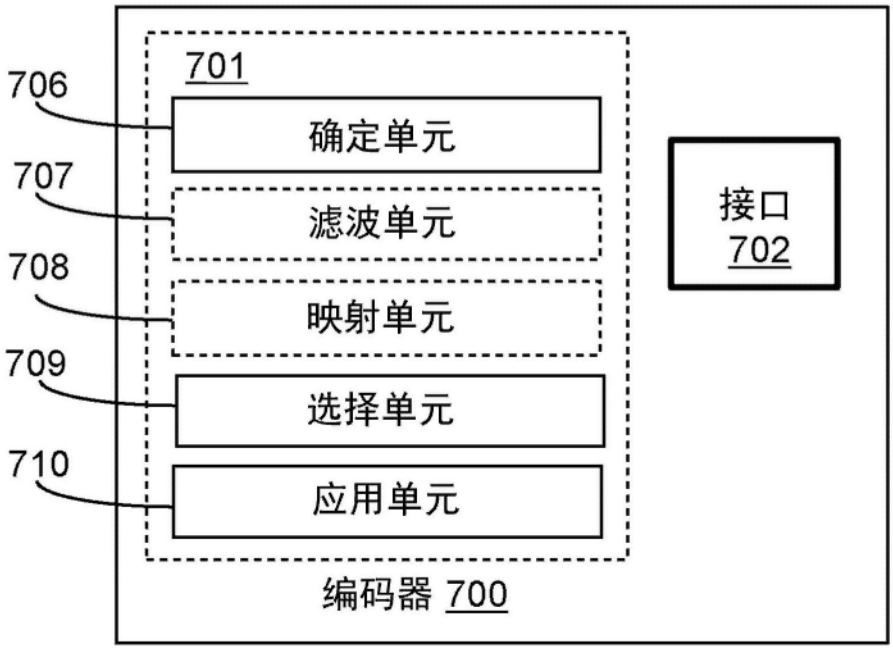

701
706
确定单元
707
滤波单元
708
映射单元
709
选择单元
710
应用单元
接口
702
编码器 700

图7c

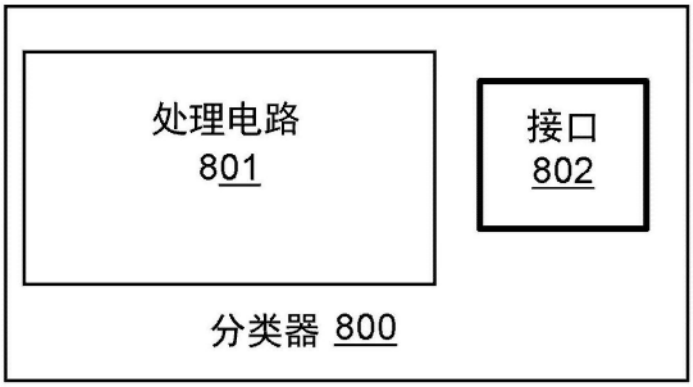

处理电路
801
接口
802
分类器 800

图8a

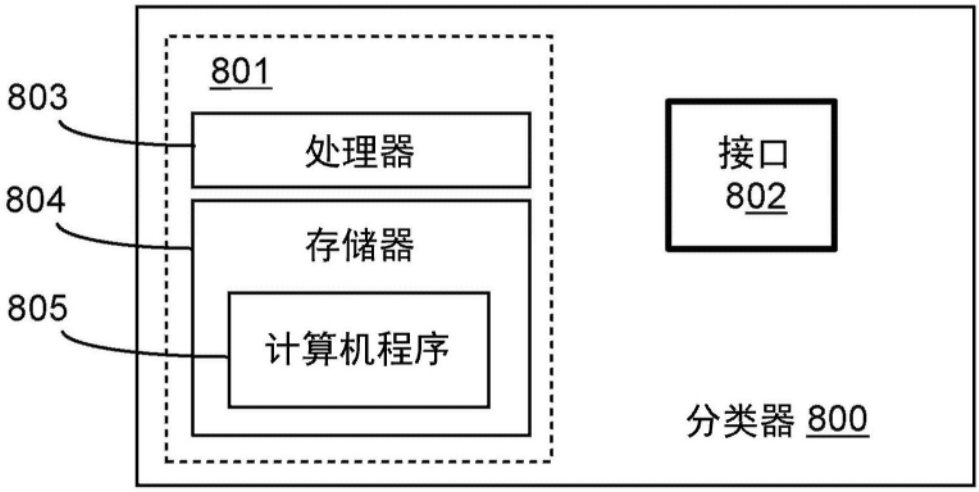

801
803
处理器
804
存储器
805
计算机程序
接口
802
分类器 800

图8b

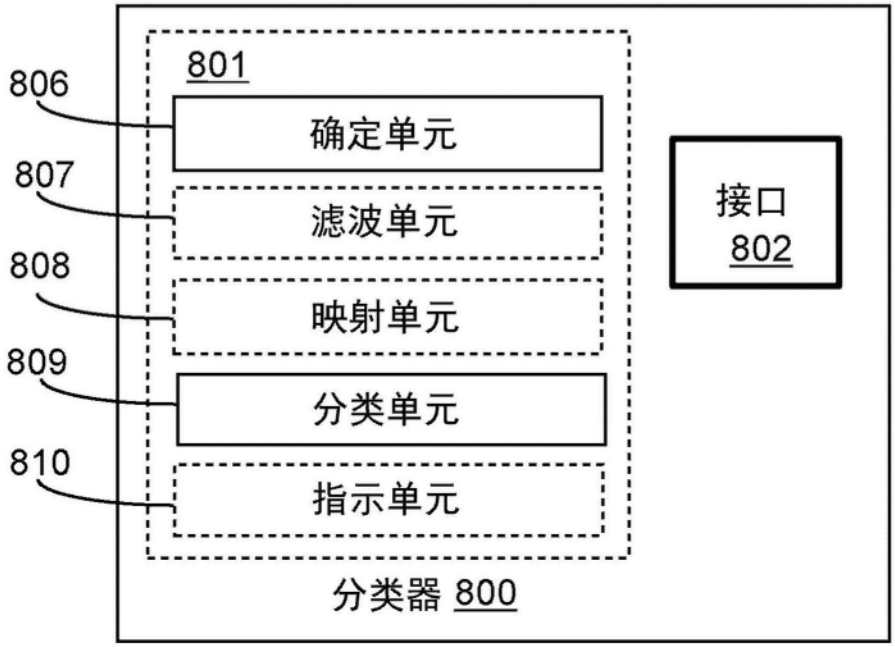

801
806
确定单元
807
滤波单元
808
映射单元
809
分类单元
810
指示单元
接口
802
分类器 800

图8c

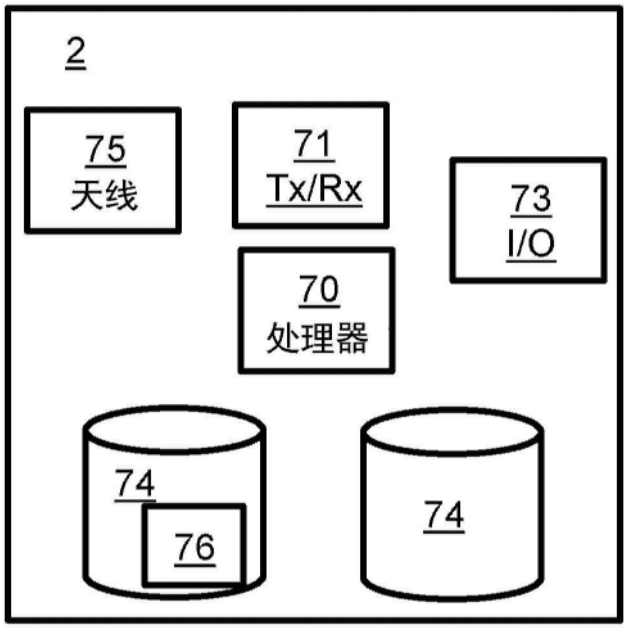

2
75
天线
71
Tx/Rx
73
I/O
70
处理器
74
76
74

图9

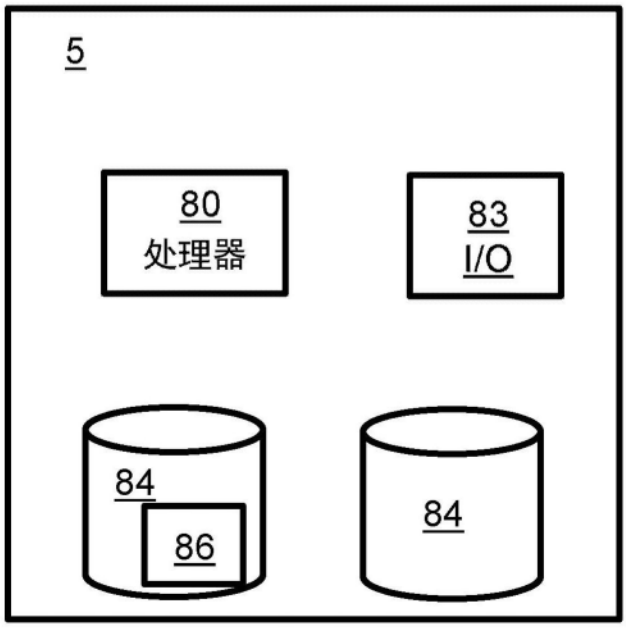


图10

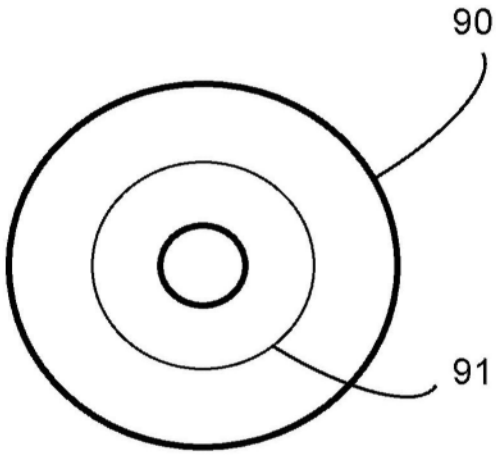


图11