



(12) 发明专利

(10) 授权公告号 CN 107958670 B

(45) 授权公告日 2021.11.19

(21) 申请号 201711421463.5

李男淑

(22) 申请日 2013.11.13

(74) 专利代理机构 北京铭硕知识产权代理有限公司 11286

(65) 同一申请的已公布的文献号  
申请公布号 CN 107958670 A

代理人 田方 曾世骁

(43) 申请公布日 2018.04.24

(51) Int.Cl.  
G10L 19/22 (2013.01)

(30) 优先权数据  
61/725,694 2012.11.13 US

(56) 对比文件  
CN 101197135 A, 2008.06.11

(62) 分案原申请数据  
201380070268.6 2013.11.13

CN 1954364 A, 2007.04.25

CN 101350199 A, 2009.01.21

(73) 专利权人 三星电子株式会社  
地址 韩国京畿道水原市

US 2012069899 A1, 2012.03.22

US 2003101050 A1, 2003.05.29

(72) 发明人 朱基岷  
安东·维克托维奇·波罗夫  
康斯坦丁·谢尔盖耶维奇·奥斯波夫

审查员 殷铸灵

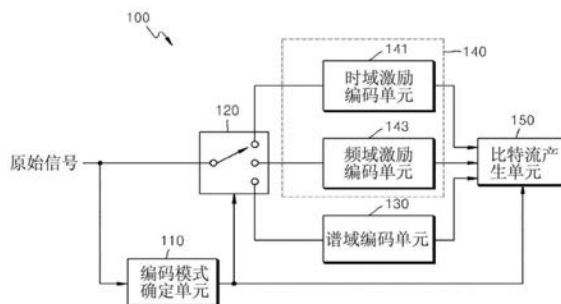
权利要求书1页 说明书12页 附图4页

(54) 发明名称

用于确定编码模式的设备以及音频编码设备

(57) 摘要

提供了一种用于确定编码模式的设备以及音频编码设备。一种确定编码模式的方法包括：根据音频信号的特性，将包括第一编码模式和第二编码模式的多个编码模式中的一个编码模式确定为初始编码模式；如果在对初始编码模式的确定中存在错误，则通过将初始编码模式校正为第三编码模式来产生经过校正的编码模式。



1. 一种用于确定编码模式的设备,所述设备包括:  
至少一个处理装置,被配置为:  
基于信号特性,在包括音乐类别和语音类别的多个类别之中确定当前帧的类别;  
从包括当前帧的多个帧获得包括第一音调、第二音调和第三音调的特征参数;  
基于所述特征参数来产生至少一个条件;  
基于所述至少一个条件,确定在所确定的当前帧的类别上是否发生错误;  
当在所确定的当前帧的类别上发生错误并且所确定的当前帧的类别是音乐类别时,将所确定的当前帧的类别校正为语音类别;  
当在所确定的当前帧的类别上发生错误并且所确定的当前帧的类别是语音类别时,将所确定的当前帧的类别校正为音乐类别,  
其中,所述第一音调、所述第二音调和所述第三音调是从不同的频段获得的。
2. 如权利要求1所述的设备,其中,所述特征参数还包括线性预测误差。
3. 如权利要求1所述的设备,其中,所述特征参数还包括浊音参数与相关度参数之间的差。
4. 一种音频编码设备,所述设备包括:  
至少一个处理装置,被配置为:  
基于信号特性,在包括音乐类别和语音类别的多个类别之中确定当前帧的类别;  
从包括当前帧的多个帧获得包括第一音调、第二音调和第三音调的特征参数;  
基于所述特征参数来产生至少一个条件;  
基于所述至少一个条件,确定在所确定的当前帧的类别上是否发生错误;  
当在所确定的当前帧的类别上发生错误并且所确定的当前帧的类别是音乐类别时,将所确定的当前帧的类别校正为语音类别;  
当在所确定的当前帧的类别上发生错误并且所确定的当前帧的类别是语音类别时,将所确定的当前帧的类别校正为音乐类别;  
基于所确定的当前帧的类别或校正后的当前帧的类别,对当前帧执行不同的编码处理,  
其中,所述第一音调、所述第二音调和所述第三音调是从不同的频段获得的。
5. 如权利要求4所述的设备,其中,所述特征参数还包括线性预测误差。
6. 如权利要求4所述的设备,其中,所述特征参数还包括浊音参数与相关度参数之间的差。

## 用于确定编码模式的设备以及音频编码设备

[0001] 本申请是向中国知识产权局提交的申请日为2013年11月13日,标题为“用于确定编码模式的方法和设备、用于对音频信号进行编码的方法和设备以及用于对音频信号进行解码的方法和设备”,申请号为201380070268.6的申请的分案申请。

### 技术领域

[0002] 与示例性实施例一致的设备和方法涉及音频编码和音频解码,更具体地讲,涉及一种通过确定适合于音频信号的特性的编码模式并防止频繁的编码模式切换来确定用于提高重构的音频信号的质量的编码模式的方法和设备,一种用于对音频信号进行编码的方法和设备以及一种用于对音频信号进行解码的方法和设备。

### 背景技术

[0003] 广为人知的是,在频域对音乐信号进行编码是有效率的并且在时域对语音信号进行编码是有效率的。因此,已提出了用于确定混合有音乐信号和语音信号的音频信号的类别并确定与所确定的类别相应的编码模式的各种技术。

[0004] 然而,由于频率编码模式切换,不仅发生延迟,还使解码的声音质量降低。此外,由于不存在用于校正最初确定的编码模式(即,类别)的技术,因此,如果在确定编码模式期间发生错误,则重构的音频信号的质量降低。

### 发明内容

[0005] 技术问题

[0006] 一个或更多个示例性实施例的多个方面提供了一种用于通过确定适合于音频信号的特性的编码模式来确定用于提高重构的音频信号的质量的编码模式的方法和设备,一种用于对音频信号进行编码的方法和设备以及一种用于对音频信号进行解码的方法和设备。

[0007] 一个或更多个示例性实施例的多个方面提供了一种用于确定适合于音频信号的特性的编码模式并减少由于频繁的编码模式切换而引起的时延的方法和设备,一种用于对音频信号进行编码的方法和设备以及一种用于对音频信号进行解码的方法和设备。

[0008] 解决方案

[0009] 根据一个或更多个示例性实施例的一方面,一种确定编码模式的方法,所述方法包括:根据音频信号的特性,将包括第一编码模式和第二编码模式的多个编码模式之中的一个编码模式确定为初始编码模式;如果在对初始编码模式的确定中存在错误,则通过将初始编码模式校正为第三编码模式来产生经过校正的编码模式。

[0010] 根据一个或更多个示例性实施例的一方面,一种对音频信号进行编码的方法,所述方法包括:根据音频信号的特性,将包括第一编码模式和第二编码模式的多个编码模式之中的一个编码模式确定为初始编码模式;如果在对初始编码模式的确定中存在错误,则通过将初始编码模式校正为第三编码模式来产生经过校正的编码模式;基于初始编码模式

或经过校正的编码模式对音频信号执行不同的编码处理。

[0011] 根据一个或多个示例性实施例的一方面,一种对音频信号进行解码的方法,所述方法包括:对包括初始编码模式和第三编码模式之一的比特流进行解析,并基于初始编码模式或第三编码模式对所述比特流执行不同的解码处理,其中,所述初始编码模式是通过根据音频信号的特性在包括第一编码模式和第二编码模式的多个编码模式之中确定一个编码模式而获得的,所述第三编码模式是在对初始编码模式的确定中存在错误的情况下从初始编码模式进行校正而得到的。

[0012] 有益效果

[0013] 根据示例性实施例,通过基于对初始编码模式的校正以及与拖尾长度相应的帧的编码模式来确定当前帧的最终编码模式,可在防止多个帧之间的频繁的编码模式切换的同时选择出适应于音频信号的特性的编码模式。

## 附图说明

[0014] 图1是示出根据示例性实施例的音频编码设备的配置的框图;

[0015] 图2是示出根据另一示例性实施例的音频编码设备的配置的框图;

[0016] 图3是示出根据示例性实施例的编码模式确定单元的配置的框图;

[0017] 图4是示出根据示例性实施例的初始编码模式确定单元的配置的框图;

[0018] 图5是示出根据示例性实施例的特征参数提取单元的配置的框图;

[0019] 图6是示出根据示例性实施例的线性预测域编码和谱域之间的自适应切换方法的示图;

[0020] 图7是示出根据示例性实施例的编码模式校正单元的操作的示图;

[0021] 图8是示出根据示例性实施例的音频解码设备的配置的框图;

[0022] 图9是示出根据另一示例性实施例的音频解码设备的配置的框图。

## 具体实施方式

[0023] 现在将详细描述实施例,其示例在附图中被示出,其中,相同的标号始终是指相同的元件。在这一点上,本实施例可具有不同的形式并且不应该被解释为受限于在此阐述的描述。因此,通过参照附图,在下面实施例仅被描述用于解释本说明书的多个方面。

[0024] 诸如“连接的”和“链接的”的术语可被用于指示直接连接或链接的状态,但应理解,另一组件可被置于其间。

[0025] 诸如“第一”和“第二”的术语可被用于描述各种组件,但所述组件不应受限于所述术语。所述术语可仅被用于使一个组件与另一组件区分开。

[0026] 在示例性实施例中描述的单元被独立示出以指示不同的特性功能,并且它不意味着每个单元由一个单独的硬件组件或软件组件形成。为了便于解释而示出每个单元,并且多个单元可形成一个单元,一个单元可被划分为多个单元。

[0027] 图1是示出根据示例性实施例的音频编码设备100的配置的框图。

[0028] 图1中示出的音频编码设备100可包括编码模式确定单元110、切换单元120、谱域编码单元130、线性预测域编码单元140和比特流产生单元150。线性预测域编码单元140可包括时域激励编码单元141和频域激励编码单元143,其中,线性预测域编码单元140可被实

现为时域激励编码单元141和频域激励编码单元143中的至少一个。除非必须被实现为单独的硬件,否则上述组件可被集成为至少一个模块并且可被实现为至少一个处理器(未示出)。这里,术语音频信号可指音乐信号、语音信号或它们的混合信号。

[0029] 参照图1,编码模式确定单元110可分析音频信号的特性以确定音频信号的类别,并根据分类的结果来确定编码模式。对编码模式的确定可以以超帧、帧或频段为单位来执行。可选择地,对编码模式的确定可以以多个超帧组、多个帧组或多个频段组为单位来执行。这里,编码模式的示例可包括谱域和时域或线性预测域,但不限于此。如果处理器的性能和处理速度足够并且由于编码模式切换引起的时延可被解决,则编码模式可被细分,并且编码方案也可根据编码模式被细分。根据示例性实施例,编码模式确定单元110可将音频信号的初始编码模式确定为谱域编码模式和时域编码模式之一。根据另一示例性实施例,编码模式确定单元110可将音频信号的初始编码模式确定为谱域编码模式、时域激励编码模式和频域激励编码模式之一。如果谱域编码模式被确定为初始编码模式,则编码模式确定单元110可将初始编码模式校正为谱域编码模式和频域激励编码模式之一。如果时域编码模式(即,时域激励编码模式)被确定为初始编码模式,则编码模式确定单元110可将初始编码模式校正为时域激励编码模式和频域激励编码模式之一。如果时域激励编码模式被确定为初始编码模式,则对最终编码模式的确定可被选择性地执行。换句话说,初始编码模式(即,时域激励编码模式)可被保持。编码模式确定单元110可确定与拖尾长度(hangover length)相应的多个帧的编码模式,并可为当前帧确定最终编码模式。根据示例性实施例,如果当前帧的初始编码模式或经过校正的编码模式与多个先前帧(例如,7个先前帧)的编码模式相同,则相应的初始编码模式或经过校正的编码模式可被确定为当前帧的最终编码模式。同时,如果当前帧的初始编码模式或经过校正的编码模式与多个先前帧(例如,7个先前帧)的编码模式不相同,则编码模式确定单元110可将恰在当前帧之前的帧的编码模式确定为当前帧的最终编码模式。

[0030] 如上所述,通过基于对初始编码模式的校正以及与拖尾长度相应的帧的编码模式来确定当前帧的最终编码模式,可在防止帧之间的频繁的编码模式切换的同时选择出适应于音频信号的特性的编码模式。

[0031] 一般来说,时域编码(即,时域激励编码)对于语音信号会是有效率的,谱域编码对于音乐信号会是有效率的,并且频域激励编码对于言语(vocal)信号和/或谐波信号会是有效率的。

[0032] 根据由编码模式确定单元110确定的编码模式,切换单元120可向谱域编码单元130或线性预测域编码单元140提供音频信号。如果线性预测域编码单元140被实现为时域激励编码单元141,则切换单元120可包括总共两个分支。如果线性预测域编码单元140被实现为时域激励编码单元141和频域激励编码单元143,则切换单元120可具有总共3个分支。

[0033] 谱域编码单元130可在谱域对音频信号进行编码。谱域可指频域或变换域。适合于谱域编码单元130的编码方法的示例可包括高级音频编码(AAC)或包括改进离散余弦变换(MDCT)和阶乘脉冲编码(FPC)的组合,但不限于此。详细地讲,其它量化技术和熵编码技术可用来代替FPC。在谱域编码单元130中对音乐信号进行编码会是有效率的。

[0034] 线性预测域编码单元140可在线性预测域对音频信号进行编码。线性预测域可指激励域或时域。线性预测域编码单元140可被实现为时域激励编码单元141,或者可被实现

为包括时域激励编码单元141和频域激励编码单元143。适合于时域激励编码单元141的编码方法的示例可包括码激励线性预测 (CELP) 或代数CELP (ACELP), 但不限于此。适合于频域激励编码单元143的编码方法的示例可包括通用信号编码 (GSC) 或变换码激励 (TCX), 但不限于此。在时域激励编码单元141中对语音信号进行编码会是有效率的, 而在频域激励编码单元143中对言语信号和/或谐波信号进行编码会是有效率的。

[0035] 比特流产生单元150可产生比特流来包括由编码模式确定单元110提供的编码模式、由谱域编码单元130提供的编码结果以及由线性预测域编码单元140提供的编码结果。

[0036] 图2是示出根据另一示例性实施例的音频编码设备200的配置的框图。

[0037] 图2中示出的音频编码设备200可包括公共预处理模块205、编码模式确定单元210、切换单元220、谱域编码单元230、线性预测域编码单元240和比特流产生单元250。这里, 线性预测域编码单元240可包括时域激励编码单元241和频域激励编码单元243, 线性预测域编码单元240可被实现为时域激励编码单元或频域激励编码单元243。与图1中示出的音频编码设备100相比, 音频编码设备200还可包括公共预处理模块205, 因此, 与音频编码设备100的组件相同的组件的描述将被省略。

[0038] 参照图2, 公共预处理模块205可执行联合立体声处理、环绕处理和/或带宽扩展处理。联合立体声处理、环绕处理和带宽扩展处理可与由特定标准 (例如, MPEG标准) 采用的那些处理相同, 但不限于此。公共预处理模块205的输出可以是在单声道、立体声声道或多声道中。根据由公共预处理模块205输出的信号的声道的数量, 切换单元220可包括至少一个开关。例如, 如果公共预处理模块205输出两个或更多个声道 (即, 立体声声道或多声道) 的信号, 则与各个声道相应的开关可被布置。例如, 立体声信号的第一声道可以是语音声道, 立体声信号的第二声道可以是音乐声道。在这种情况下, 音频信号可被同时提供给两个开关。由公共预处理模块205产生的附加信息可被提供给比特流产生单元250并被包括在比特流中。所述附加信息对于在解码端执行联合立体声处理、环绕处理和/或带宽扩展处理是必要的, 并且可包括空间参数、包络信息、能量信息等。然而, 基于所应用的处理技术, 可存在各种附加信息。

[0039] 根据示例性实施例, 在公共预处理模块205, 可基于编码域而不同地执行带宽扩展处理。核心频段中的音频信号可通过使用时域激励编码模式或频域激励编码模式来处理, 而带宽扩展频段中的音频信号可在时域中被处理。时域中的带宽扩展处理可包括多个模式 (包括浊音模式或清音模式)。可选择地, 核心频段中的音频信号可通过使用谱域编码模式来处理, 而带宽扩展频段中的音频信号可在频域中被处理。频域中的带宽扩展处理可包括多个模式 (包括瞬变模式、一般模式或谐波模式)。为了在不同域中执行带宽扩展处理, 由编码模式确定单元110确定的编码模式可作为信令信息被提供给公共预处理模块205。根据示例性实施例, 核心频段的最后部分和带宽扩展频段的开始部分可能在一定程度上彼此重叠。重叠部分的位置和尺寸可被预先设置。

[0040] 图3是示出根据示例性实施例的编码模式确定单元300的配置的框图。

[0041] 图3中示出的编码模式确定单元300可包括初始编码模式确定单元310和编码模式校正单元330。

[0042] 参照图3, 初始编码模式确定单元310可通过使用从音频信号提取出的特征参数来确定音频信号是音乐信号还是语音信号。如果音频信号被确定为语音信号, 则线性预测域

编码会是合适的。同时,如果音频信号被确定为音乐信号,则谱域编码会是适合的。初始编码模式确定单元310可通过使用从音频信号提取出的特征参数来确定音频信号的类别,其中,音频信号的类别指示是谱域编码、时域激励编码还是频域激励编码适合于该音频信号。可基于音频信号的类别来确定相应编码模式。如果(图1的)切换单元(120)具有两个分支,则编码模式可以以1比特来表示。如果(图1的)切换单元(120)具有三个分支,则编码模式可以以2比特来表示。初始编码模式确定单元310可通过使用现有技术中已知的各种技术中的任意技术来确定音频信号是音乐信号还是语音信号。其示例可包括USAC标准的编码器部分中公开的FD/LPD分类或ACELP/TCX分类以及AMR标准中使用的ACELP/TCX分类,但不限于此。换句话说,可通过使用除在此描述的根据实施例的方法以外的各种任意方法来确定初始编码模式。

[0043] 编码模式校正单元330可通过使用校正参数对由初始编码模式确定单元310确定的初始编码模式进行校正来确定经过校正的编码模式。根据示例性实施例,如果谱域编码模式被确定为初始编码模式,则基于校正参数,初始编码模式可被校正为频域激励编码模式。如果时域编码模式被确定为初始编码模式,则基于校正参数,初始编码模式可被校正为频域激励编码模式。换句话说,通过使用校正参数,确定在对初始编码模式的确定中是否存在错误。如果确定在对初始编码模式的确定中不存在错误,则初始编码模式可被保持。相反,如果确定在对初始编码模式的确定中存在错误,则初始编码模式可被校正。可获得从谱域编码模式到频域激励编码模式以及从时域激励编码模式到频域激励编码模式的对初始编码模式的校正。

[0044] 同时,初始编码模式或经过校正的编码模式可以是用于当前帧的临时编码模式,其中,可将用于当前帧的临时编码模式与用于预设拖尾长度内的先前帧的编码模式进行比较,并可确定用于当前帧的最终编码模式。

[0045] 图4是示出根据示例性实施例的初始编码模式确定单元400的配置的框图。

[0046] 图4中示出的初始编码模式确定单元400可包括特征参数提取单元410和确定单元430。

[0047] 参照图4,特征参数提取单元410可从音频信号提取用于确定编码模式的所必要的特征参数。提取的特征参数的示例包括音高(pitch)参数、浊音参数、相关度参数和线性预测误差之中的至少一个或两个,但不限于此。以下将给出对各个参数的详细描述。

[0048] 首先,第一特征参数 $F_1$ 与音高参数有关,其中,可通过使用在当前帧和至少一个先前帧中检测到的 $N$ 个音高值来确定音高的表现。为了防止效果随机偏离或防止错误的音高值,可去除与所述 $N$ 个音高值的平均值明显不同的 $M$ 个音高值。这里, $N$ 和 $M$ 可以是预先经由实验或仿真而被获取的值。此外, $N$ 可被预先设置,并且将被移除的音高值与所述 $N$ 个音高值之间的平均值之间的差可预先经由实验或仿真而被确定。通过使用关于 $(N-M)$ 个音高值的均值 $m_p'$ 和方差 $\sigma_p'$ ,第一特征参数 $F_1$ 可如下面的等式1中所示被表达。

[0049] [等式1]

$$[0050] \quad F_1 = \frac{\sigma_p'}{m_p'}$$

[0051] 第二特征参数 $F_2$ 也与音高参数有关,并可指示在当前帧中检测到的音高值的可靠性。通过使用在当前帧的两个子帧 $SF_1$ 和 $SF_2$ 中分别检测到的音高值的方差 $\sigma_{SF1}$ 和 $\sigma_{SF2}$ ,第二特

征参数 $F_2$ 可如下面的等式2中所示被表达。

[0052] [等式2]

$$[0053] \quad F_2 = \frac{\text{cov}(SF_1, SF_2)}{\sigma_{SF_1} \sigma_{SF_2}}$$

[0054] 这里,  $\text{cov}(SF_1, SF_2)$  表示子帧 $SF_1$ 和子帧 $SF_2$ 之间的协方差。换句话说, 第二特征参数 $F_2$ 将两个子帧之间的相关度指示为音高距离。根据示例性实施例, 当前帧可包括两个或更多个子帧, 等式2可基于子帧的数量而被修改。

[0055] 基于浊音参数Voicing和相关度参数Corr, 第三特征参数 $F_3$ 可如下面的等式3中所示被表达。

[0056] [等式3]

$$[0057] \quad F_3 = \sqrt{Q \frac{C \text{Voicing} - \text{Corr} C^2}{N}}$$

[0058] 这里, 浊音参数Voicing与声音的言语特征相关, 并且可通过现有技术中已知的各种方法中的任意方法来获取, 而相关度参数Corr可通过对针对每个频段的帧之间的相关度求和来获取。

[0059] 第四特征参数 $F_4$ 与线性预测误差 $E_{LPC}$ 相关并可如下面的等式4中所示被表达。

[0060] [等式4]

$$[0061] \quad F_4 = \frac{\sqrt{(E_{LPCi} - M(E_{LPC}))^2}}{N}$$

[0062] 这里,  $M(E_{LPC})$  表示N个线性预测误差的平均值。

[0063] 确定单元430可通过使用由特征参数提取单元410提供的至少一个特征参数来确定音频信号的类别, 并可基于所确定的类别来确定初始编码模式。确定单元430可采用软判决机制, 其中, 在软判决机制中, 可根据每个特征参数形成至少一个混合。根据示例性实施例, 可通过基于混合 (mixture) 概率使用高斯混合模型 (GMM) 来确定音频信号的类别。关于一个混合的概率 $f(x)$ 可根据下面的等式5来计算。

[0064] [等式5]

$$[0065] \quad f(x) = \frac{1}{\sqrt{(2\pi)^N \det(C^{-1})}} e^{-0.5(x-m)^T C^{-1}(x-m)}$$

[0066]  $x = (x_1, \dots, x_N)$

[0067]  $m = (C_{x_1 C}, \dots, C_{x_N C})$

[0068] 这里,  $x$ 表示特征参数的输入矢量,  $m$ 表示混合,  $c$ 表示协方差矩阵。

[0069] 确定单元430可通过使用下面的等式6来计算音乐概率 $P_m$ 和语音概率 $P_s$ 。

[0070] [等式6]

$$[0071] \quad P_m = \prod_{i \in M} p_i, P_s = \prod_{i \in S} p_i$$

[0072] 这里, 可通过将与适合于音乐确定的特征参数相关的M个混合的概率 $P_i$ 相加来



计算音乐概率 $P_m$ ,而可通过将与适合用于语音确定的特征参数相关的 $S$ 个混合的概率 $P_i$ 相加来计算语音概率 $P_s$ 。

[0073] 同时,为了提高精确度,可根据下面的等式7来计算音乐概率 $P_m$ 和语音概率 $P_s$ 。

[0074] [等式7]

$$[0075] \quad P_m = \sum_{i \in M} p_i(1-p_i^{err}) + \sum_{i \in S} p_i(p_i^{err})$$

$$[0076] \quad P_s = \sum_{i \in S} p_i(1-p_i^{err}) + \sum_{i \in M} p_i(p_i^{err})$$

[0077] 这里,  $p_i^{err}$  表示每个混合的误差概率。可通过使用每个混合对包括纯净语音信号和纯净音乐信号的训练数据进行分类并对错误分类的数量进行计数来获取所述误差概率。

[0078] 接下来,可根据下面的等式8,针对与恒定拖尾长度相同数量的多个帧,计算所有帧仅包括音乐信号的音乐概率 $P^M$ 和所有帧仅包括语音信号的语音概率 $P^S$ 。拖尾长度可被设置为8,但不限于此。八个帧可包括当前帧和7个先前帧。

[0079] [等式8]

$$[0080] \quad P^M = \frac{\sum_{i=0}^{-7} P_m^{(i)}}{\sum_{i=0}^{-7} P_m^{(i)} + \sum_{i=0}^{-7} P_s^{(i)}}$$

$$[0081] \quad P^S = \frac{\sum_{i=0}^{-7} P_s^{(i)}}{\sum_{i=0}^{-7} P_m^{(i)} + \sum_{i=0}^{-7} P_s^{(i)}}$$

[0082] 接下来,可通过使用利用等式5或等式6获取的音乐概率 $P_m$ 或语音概率 $P_s$ 来计算多个状况(condition)集合 $\{D_i^M\}$ 和 $\{D_i^S\}$ 。下面将参照图6给出其详细描述。这里,可按照每个状况对于音乐具有值1并且对于语音具有值0的方式进行设置。

[0083] 参照图6,在操作610和操作620,可从通过使用音乐概率 $P_m$ 和语音概率 $P_s$ 计算出的多个状况集合 $\{D_i^M\}$ 和 $\{D_i^S\}$ 来获取音乐状况之和 $M$ 以及语音状况之和 $S$ 。换句话说,音乐状况之和 $M$ 以及语音状况之和 $S$ 可如下面的等式9中所示来表达。

[0084] [等式9]

$$[0085] \quad M = \sum_i D_i^M$$

$$[0086] \quad S = \sum_i D_i^S$$

[0087] 在操作630,将音乐状况之和 $M$ 与指定的阈值 $T_m$ 进行比较。如果音乐状况之和 $M$ 大于所述阈值 $T_m$ ,则当前帧的编码模式被切换为音乐模式(即,谱域编码模式)。如果音乐状况之和 $M$ 小于或等于阈值 $T_m$ ,则当前帧的编码模式不被改变。

[0088] 在操作640,将语音状况之和 $S$ 与指定阈值 $T_s$ 进行比较。如果语音状况之和 $S$ 大于阈

值 $T_s$ ,则当前帧的编码模式被切换为语音模式(即,线性预测域编码模式)。如果语音状况之和 $S$ 小于或等于阈值 $T_s$ ,则当前帧的编码模式不被改变。

[0089] 阈值 $T_m$ 和阈值 $T_s$ 可被设置为预先经由实验或仿真而获取的值。

[0090] 图5是示出根据示例性实施例的特征参数提取单元500的配置的框图。

[0091] 图5中示出的初始编码模式确定单元500可包括变换单元510、频谱参数提取单元520、时间参数提取单元530和确定单元540。

[0092] 在图5中,变换单元510可将原始音频信号从时域变换到频域。这里,变换单元510可应用各种任意变换技术以将音频信号从时域表示为谱域。所述技术的实例可包括快速傅里叶变换(FFT)、离散余弦变换(DCT)或改进离散余弦变换(MDCT),但不限于此。

[0093] 频谱参数提取单元520可从由变换单元510提供的频域音频信号提取至少一个频谱参数。频谱参数可被归类为短期特征参数和长期特征参数。可从当前帧获取短期特征参数,而可从包括当前帧和至少一个先前帧的多个帧获取长期特征参数。

[0094] 时间参数提取单元530可从时域音频信号提取至少一个时间参数。时间参数也可被归类为短期特征参数和长期特征参数。可从当前帧获取短期特征参数,而可从包括当前帧和至少一个先前帧的多个帧获取长期特征参数。

[0095] (图4的)确定单元(430)可通过使用由频谱参数提取单元520提供的频谱参数以及由时间参数提取单元530提供的时间参数来确定音频信号的类别,并可基于所确定的类别来确定初始编码模式。(图4的)确定单元(430)可采用软判决机制。

[0096] 图7是示出根据示例性实施例的编码模式校正单元310的操作的示图。

[0097] 参照图7,在操作700,由初始编码模式确定单元310确定的初始编码模式被接收,并且可确定编码模式是时域模式(即,时域激励模式)还是谱域模式。

[0098] 在操作701,如果在操作700确定初始编码模式是谱域模式( $state_{TS}=1$ ),则可检查指示频域激励编码是否更加合适的索引 $state_{TSS}$ 。可通过使用不同频段的音调来获取指示频域激励编码(例如,GSC)是否更加合适的索引 $state_{TSS}$ 。下面将给出其详细描述。

[0099] 低频段信号的音调可被获取为具有包括最小值的多个较小值的多个频谱系数之和与具有针对给定频段的最大值的频谱系数之间的比率。如果给定频段是0~1kHz、1~2kHz和2~4kHz,则各个频段的音调 $t_{01}$ 、 $t_{12}$ 和 $t_{24}$ 以及低频段信号(即,核心频段)的音调 $t_L$ 可如下面的等式10中所示被表达。

[0100] [等式10]

$$[0101] \quad t_{01} = 0.2 \log_{10} \left( \frac{\max(x_i)}{\prod_{j=0}^{M-1} \text{sort}(x_j)} \right), i, j \in [0, \dots, 1kHz]$$

$$[0102] \quad t_{12} = 0.2 \log_{10} \left( \frac{\max(x_i)}{\prod_{j=0}^{M-1} \text{sort}(x_j)} \right), i, j \in [1, \dots, 2kHz]$$

$$[0103] \quad t_{24} = 0.2 \log_{10} \left( \frac{\max(x_i)}{\prod_{j=0}^{M-1} \text{sort}(x_j)} \right), i, j \in [2, \dots, 4kHz]$$

[0104]  $t_L = \max(t_{01}, t_{12}, t_{24})$

[0105] 同时,线性预测误差可通过使用线性预测编码(LPC)滤波器来获取并可被用于去除强音调分量。换句话说针对强音调分量,谱域编码模式比频域激励编码模式更加有效。

[0106] 用于通过使用如上所述获取的音调和线性预测误差切换到频域激励编码模式的前置条件 $\text{cond}_{\text{front}}$ 可如下面的等式11中所示被表达。

[0107] [等式11]

[0108]  $\text{cond}_{\text{front}} = t_{12} > t_{12\text{front}} \text{ 且 } t_{24} > t_{24\text{front}} \text{ 且 } t_L > t_{L\text{front}} \text{ 且 } \text{err} > \text{err}_{\text{front}}$

[0109] 这里, $t_{12\text{front}}$ 、 $t_{24\text{front}}$ 、 $t_{L\text{front}}$ 和 $\text{err}_{\text{front}}$ 是阈值,并可具有预先经由实验或仿真而获取的值。

[0110] 同时,用于通过使用如上所述获取的音调和线性预测误差来完成频域激励编码模式的后置条件 $\text{cond}_{\text{back}}$ 可如下面的等式12中所示被表达。

[0111] [等式12]

[0112]  $\text{cond}_{\text{back}} = t_{12} < t_{12\text{back}} \text{ 且 } t_{24} < t_{24\text{back}} \text{ 且 } t_L < t_{L\text{back}}$

[0113] 这里, $t_{12\text{back}}$ 、 $t_{24\text{back}}$ 、 $t_{L\text{back}}$ 是阈值并可具有预先经由实验或仿真而获取的值。

[0114] 换句话说,可通过确定等式11中所示的前置条件是否被满足或是等式12中所示的后置条件是否被满足来确定索引 $\text{state}_{\text{TTSS}}$ 是否为1,其中,索引 $\text{state}_{\text{TTSS}}$ 指示频域激励编码(例如,GSC)是否比谱域编码更加合适。这里,对等式12中示出的后置条件的确定可以是可选的。

[0115] 在操作702,如果索引 $\text{state}_{\text{TTSS}}$ 是1,则频域激励编码模式可被确定为最终编码模式。在这种情况下,作为初始编码模式的谱域编码模式被校正为作为最终编码模式的频域激励编码模式。

[0116] 在操作705,如果在操作701确定索引 $\text{state}_{\text{TTSS}}$ 是0,则可检查用于确定音频信号是否包括强语音特性的索引 $\text{state}_{\text{SS}}$ 。如果在对谱域编码模式的确定中存在错误,则频域激励编码模式会比谱域编码模式更加有效。可通过使用浊音参数和相关度参数之间的差 $\text{vc}$ 来获取用于确定音频信号是否包括强语音特性的索引 $\text{state}_{\text{SS}}$ 。

[0117] 用于通过使用浊音参数和相关度参数之间的差 $\text{vc}$ 来切换到强语音模式的前置条件 $\text{cond}_{\text{front}}$ 可如下面的等式13中所示被表达。

[0118] [等式13]

[0119]  $\text{cond}_{\text{front}} = \text{vc} > \text{vc}_{\text{front}}$

[0120] 这里, $\text{vc}_{\text{front}}$ 是阈值并可具有预先经由实验或仿真而获取的值。

[0121] 同时,用于通过使用浊音参数和相关度参数之间的差 $\text{vc}$ 来结束强语音模式的后置条件 $\text{cond}_{\text{back}}$ 可如下面的等式14中所示被表达。

[0122] [等式14]

[0123]  $\text{cond}_{\text{back}} = \text{vc} < \text{vc}_{\text{back}}$

[0124] 这里, $\text{vc}_{\text{back}}$ 是阈值并可具有预先经由实验或仿真而获取的值。

[0125] 换句话说,在操作705,可通过确定等式13中示出的前置条件是否被满足或是等式14中示出的后置条件是否未被满足来确定索引 $\text{state}_{\text{SS}}$ 是否为1,其中,索引 $\text{state}_{\text{SS}}$ 指示频域激励编码(例如,GSC)是否比谱域编码更加合适。这里,对等式14中示出的对后置条件的确定可以是可选的。

[0126] 在操作706,如果在操作705确定索引 $state_{SS}$ 为0(即,音频信号不包括强语音特性),则谱域编码模式可被确定为最终编码模式。在这种情况下,作为初始编码模式的谱域编码模式被保持为最终编码模式。

[0127] 在操作707,如果在操作705确定索引 $state_{SS}$ 为1(即,音频信号包括强语音特性),则频域激励编码模式可被确定为最终编码模式。在这种情况下,作为初始编码模式的谱域编码模式被校正为作为最终编码模式的频域激励编码模式。

[0128] 通过执行操作700、701和705,对作为初始编码模式的谱域编码模式的确定中的错误可被校正。详细地讲,作为初始编码模式的谱域编码模式可被保持作为最终编码模式,或可被切换为频域激励编码模式作为最终编码模式。

[0129] 同时,如果在操作700确定初始编码模式是线性预测域编码模式( $state_{TS} = 0$ ),则用于确定音频信号是否包括强音乐特性的索引 $state_{SM}$ 可被检查。如果在对线性预测域编码模式(即,时域激励编码模式)的确定中存在错误,则频域激励编码模式可能比时域激励编码模式更加有效。可通过使用从1减去浊音参数和相关度参数之间的差 $vc$ 而获取的值 $1-vc$ 来获取用于确定音频信号是否包括强音乐特性的 $state_{SM}$ 。

[0130] 用于通过使用通过从1减去浊音参数和相关度参数之间的差 $vc$ 而获取的值 $1-vc$ 而切换到强音乐模式的前置条件 $cond_{front}$ 可如下面的等式15中所示被表达。

[0131] [等式15]

$$[0132] \quad cond_{front} = 1 - vc > vcm_{front}$$

[0133] 这里, $vcm_{front}$ 是阈值并可具有预先经由实验或仿真而获取的值。

[0134] 同时,用于通过使用通过从1减去浊音参数和相关度参数之间的差 $vc$ 而获取的值 $1-vc$ 而结束强音乐模式的后置条件 $cond_{back}$ 可如下面的等式16中所示被表达。

[0135] [等式16]

$$[0136] \quad cond_{back} = 1 - vc < vcm_{back}$$

[0137] 这里, $vcm_{back}$ 是阈值并可具有预先经由实验或仿真而获取的值。

[0138] 换句话说,在操作709,可通过确定等式15中示出的前置条件是否被满足或是等式16中示出的后置条件是否未被满足来确定索引 $state_{SM}$ 是否为1,其中,索引 $state_{SM}$ 指示频域激励编码(例如,GSC)是否比时域激励编码更适合。这里,对等式16中示出的后置条件的确定可以是可选的。

[0139] 在操作710,如果在操作709确定索引 $state_{SM}$ 为0(即,音频信号不包括强音乐特性),则时域激励编码模式可被确定为最终编码模式。在这种情况下,作为初始编码模式的线性预测域编码模式被切换为作为最终编码模式的时域激励编码模式。根据示例性实施例,如果线性预测域编码模式与时域激励编码模式对应,则可考虑初始编码模式保持不变。

[0140] 在操作707,如果在操作709确定索引 $state_{SM}$ 为1(即,音频信号包括强音乐特性),则频域激励编码模式可被确定为最终编码模式。在这种情况下,作为初始编码模式的线性预测域编码模式被校正为作为最终编码模式的频域激励编码模式。

[0141] 通过执行操作700和709,对初始编码模式的确定中的错误可被校正。详细地讲,作为初始编码模式的线性预测域编码模式(例如,时域激励编码模式)可被保持作为最终编码模式,或者可被切换为频域激励编码模式作为最终编码模式。

[0142] 根据示例性实施例,用于确定音频信号是否包括强音乐特性以校正对线性预测域

编码模式的确定中的错误的操作709可以是可选的。

[0143] 根据另一示例性实施例,执行用于确定音频信号是否包括强语音特性的操作705以及用于确定频域激励编码模式是否适合的操作701的顺序可被颠倒。换句话说,在操作700之后,可首先执行操作705,然后可执行操作701。在这种情况下,用于进行确定的参数可按照必要的需求而被改变。

[0144] 图8是示出根据示例性实施例的音频解码设备800的配置的框图。

[0145] 图8中示出的音频解码设备800可包括比特流解析单元810、谱域解码单元820、线性预测域解码单元830和切换单元840。线性预测域解码单元830可包括时域激励解码单元831和频域激励解码单元833,其中,线性预测域解码单元830可被实现为时域激励解码单元831和频域激励解码单元833中的至少一个。除非必须被实现为单独的硬件,否则上述组件可被集成为至少一个模块,并可被实现为至少一个处理器(未示出)。

[0146] 参照图8,比特流解析单元810可对接收到的比特流进行解析并对关于编码模式和编码数据的信息进行分离。编码模式可与通过根据音频信号的特性在包括第一编码模式和第二编码模式的多个编码模式之中确定一个编码模式而获取的初始编码模式相应,或者可与在对初始编码模式的确定中存在错误的情况下从初始编码模式校正得到的第三编码模式相应。

[0147] 谱域解码单元820可对来自分离的编码数据的在谱域中被编码的数据进行解码。

[0148] 线性预测域解码单元830可对来自分离的编码数据的在线性预测域中被编码的数据进行解码。如果线性预测域解码单元830包括时域激励解码单元831和频域激励解码单元833,则线性预测域解码单元830可针对分离的编码数据执行时域激励解码或频域激励解码。

[0149] 切换单元840可对由谱域解码单元820重构的信号或由线性预测域解码单元830重构的信号进行切换,并可提供切换的信号作为最终重构的信号。

[0150] 图9是示出根据另一示例性实施例的音频解码设备900的配置的框图。

[0151] 音频解码设备900可包括比特流解析单元910、谱域解码单元920、线性预测域解码单元930、切换单元940和公共后处理模块950。线性预测域解码单元930可包括时域激励解码单元931和频域激励解码单元933,其中,线性预测域解码单元930可被实现为时域激励解码单元931和频域激励解码单元933中的至少一个。除非必须被实现为单独的硬件,否则上述组件可被集成为至少一个模块,并可被实现为至少一个处理器(未示出)。与图8中示出的音频解码设备800相比,音频解码设备900还可包括公共后处理模块950,因此,将省略对与音频解码设备800的组件相同的组件的描述。

[0152] 参照图9,公共后处理模块950可执行与(图2的)公共预处理模块(205)相应的联合立体声处理、环绕处理和/或带宽扩展处理。

[0153] 根据示例性实施例的方法可被编写为计算机可执行程序并可被实现在通用数字计算机中,其中,所述通用数字计算机通过使用非暂时性计算机可读记录介质来执行程序。此外,可在实施例中使用的数据结构、程序指令或数据文件可以以各种方式被记录在非暂时性计算机可读记录介质中。非暂时性计算机可读记录介质是可存储其后可由计算机系统读出的数据的任意数据存储装置。非暂时性计算机可读记录介质的示例包括:磁介质(诸如硬盘、软盘和磁带)、光学记录介质(诸如CD ROM盘和DVD)、磁光介质(诸如光盘)以及专门配

置为存储和执行程序指令的硬件装置(诸如ROM、RAM、闪存等)。此外,非暂时性计算机可读记录介质可以是用于传输指定程序指令、数据结构等的信号的传输介质。程序指令的示例可不仅包括由编译器产生的机器语言代码,还可包括可由计算机使用解释器等执行的高级语言代码。

[0154] 尽管在上面已具体显示和描述了示例性实施例,但是本领域的普通技术人员将理解,在不脱离权利要求所限定的本发明构思的精神和范围的情况下,可以对其进行形式和细节上的各种改变。示例性实施例应被认为仅是描述性的意义而不是为了限制的目的。因此,本发明构思的范围不是由示例性实施例的详细描述来限定,而是由权利要求来限定,并且所述范围内的所有差异将被解释为包括在本发明构思中。

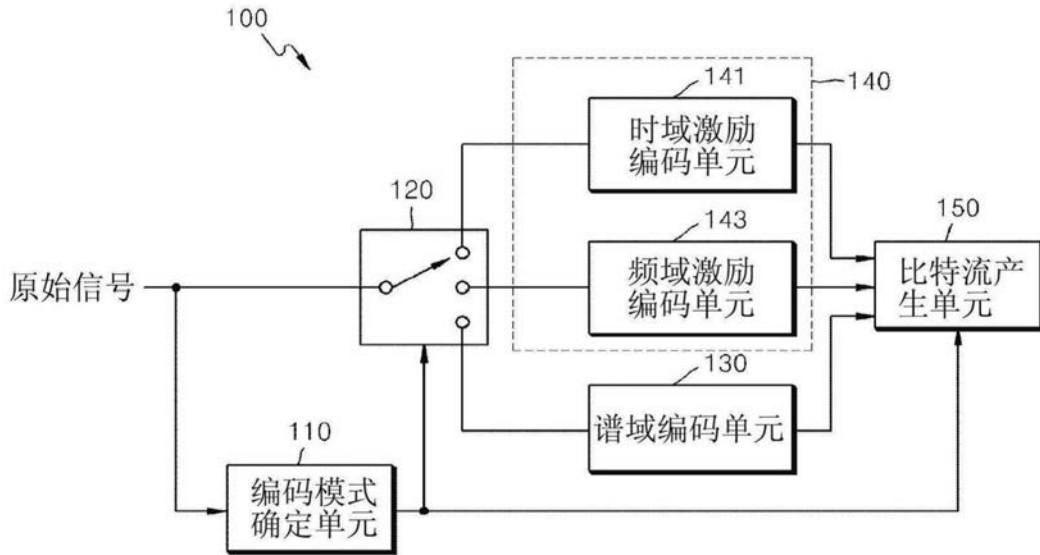


图1

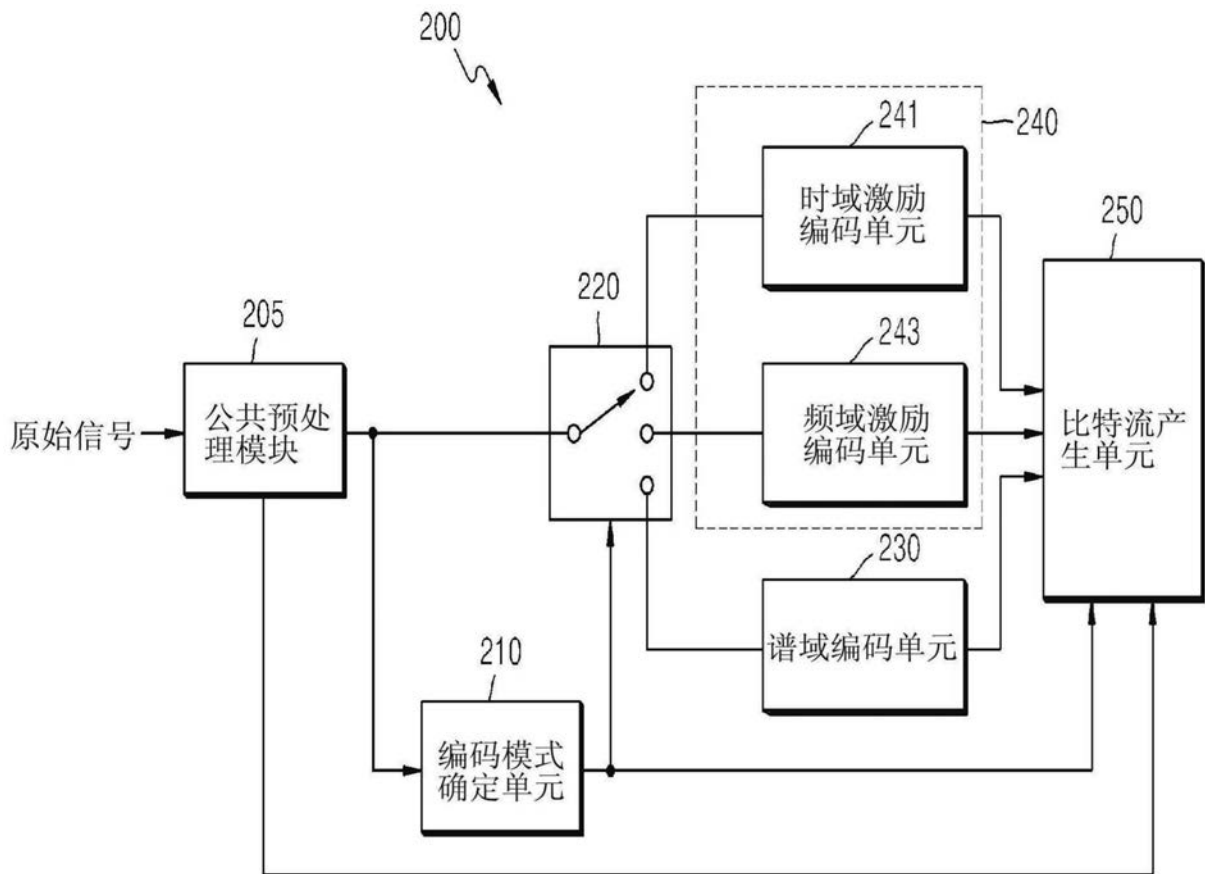


图2

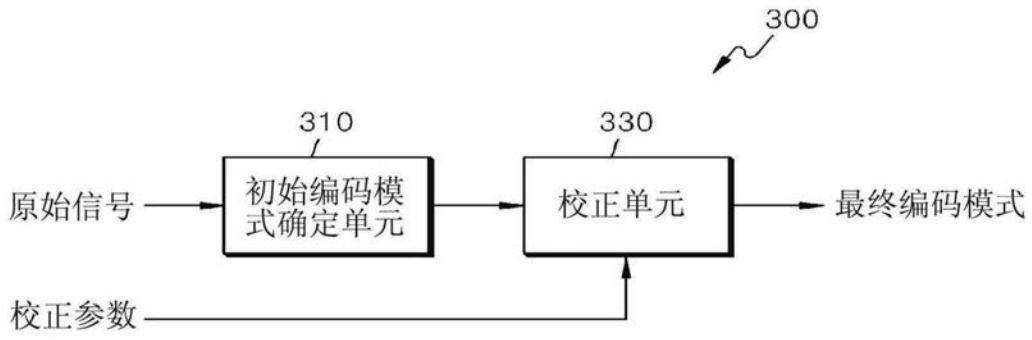


图3

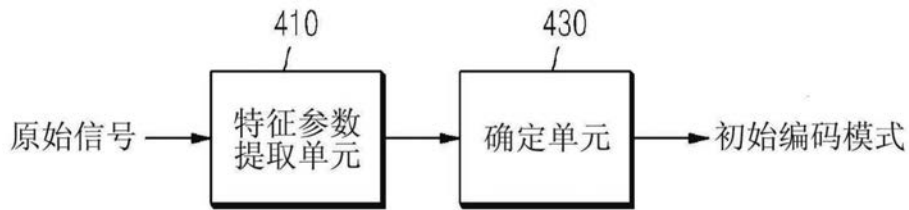


图4

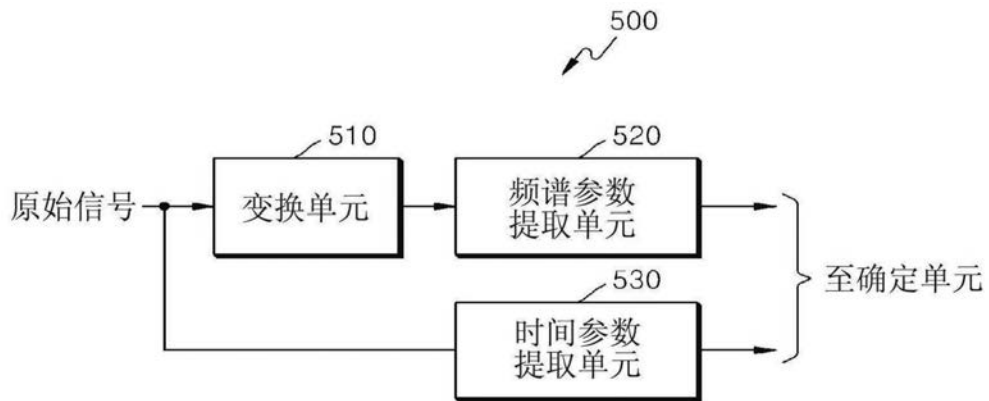


图5



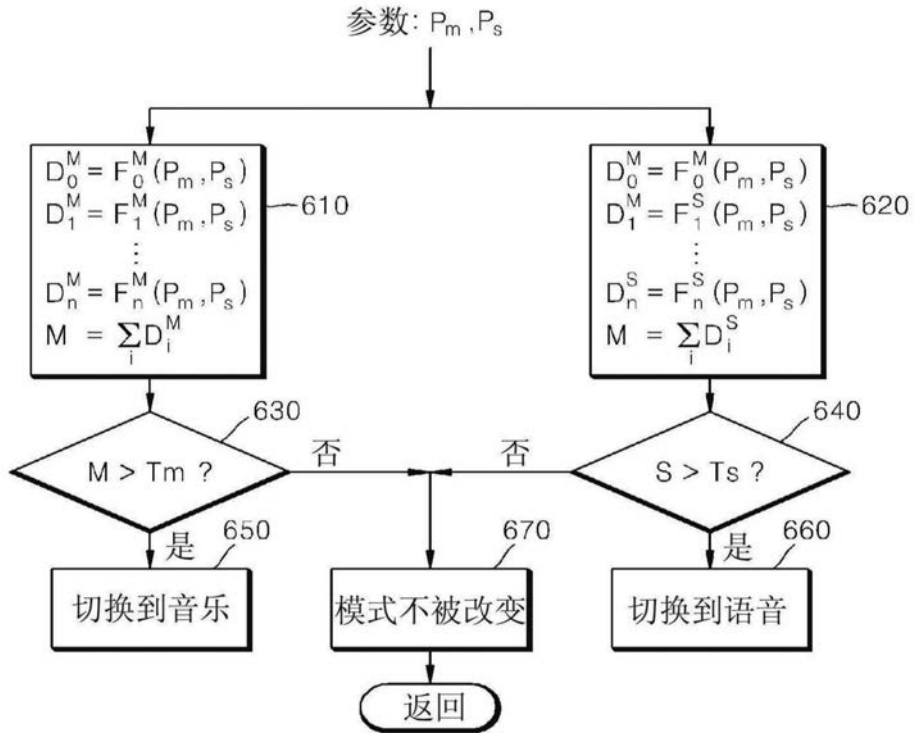


图6

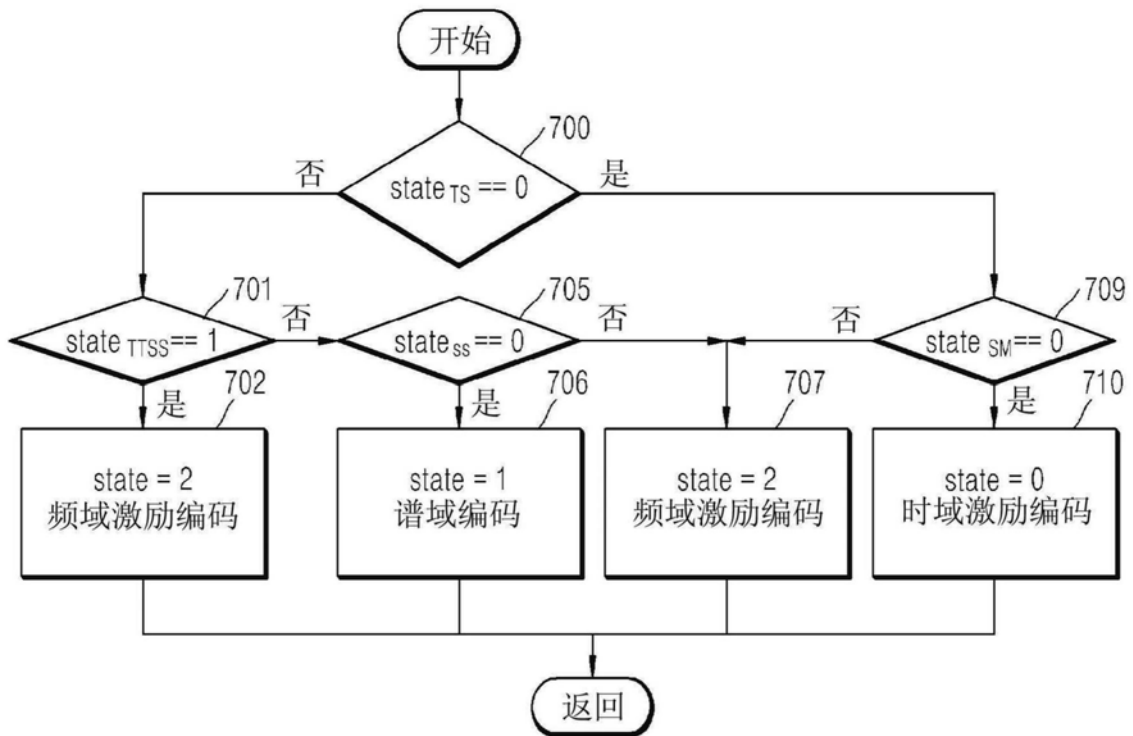


图7

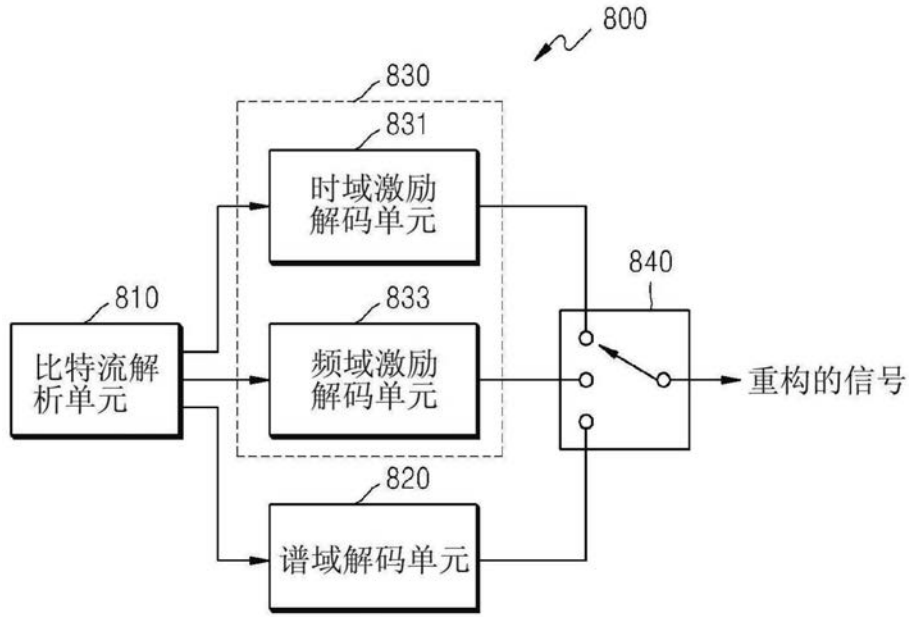


图8

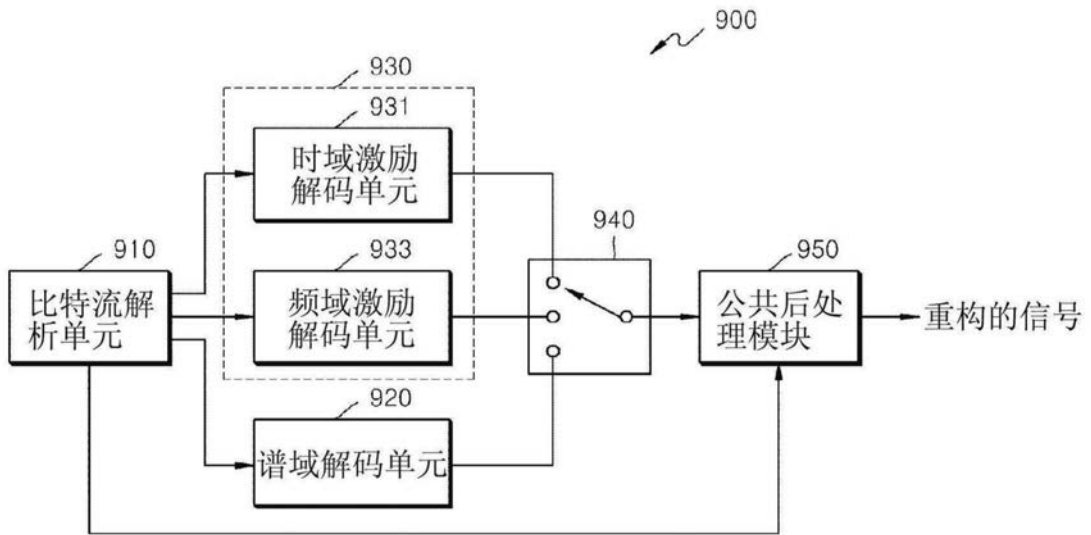


图9