



## (12)发明专利

(10)授权公告号 CN 106910509 B

(45)授权公告日 2020.08.18

(21)申请号 201710020311.8

(22)申请日 2012.11.01

(65)同一申请的已公布的文献号  
申请公布号 CN 106910509 A

(43)申请公布日 2017.06.30

(30)优先权数据  
61/555,246 2011.11.03 US(62)分案原申请数据  
201280065936.1 2012.11.01(73)专利权人 沃伊斯基公司  
地址 加拿大魁北克

(72)发明人 T.瓦兰考特 M.杰里尼克

(74)专利代理机构 北京市柳沈律师事务所  
11105  
代理人 于小宁 李文娟

(51)Int.Cl.

G10L 19/20(2013.01)

G10L 19/26(2013.01)

G10L 19/08(2013.01)

(续)

(56)对比文件

US 2004/0002856 A1, 2004.01.01

洪侃 等.《基于子带清浊音模式的声码器增益参数抗误码算法》.《清华大学学报(自然科学版)》.2008,

W.B. Kleijn et al..《A 5.85 kbits CELP algorithm for cellular applications》.《1993 IEEE International Conference on Acoustics, Speech, and Signal Processing》.1993,

B. Atal et al..《A new model of LPC excitation for producing natural-sounding speech at low bit rates》.《ICASSP '82. IEEE International Conference on Acoustics, Speech, and Signal Processing》.1982,

Daniele Giacobello et al..《Retrieving Sparse Patterns Using a Compressed Sensing Framework: Applications to Speech Coding Based on Sparse Linear Prediction》.《IEEE Signal Processing Letters ( Volume: 17 , Issue: 1 , Jan. 2010 )》.2010,

审查员 王立华

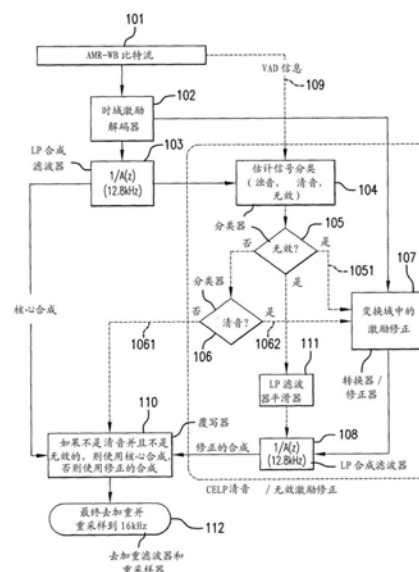
权利要求书3页 说明书12页 附图4页

(54)发明名称

用于修正通用音频合成的设备及其方法

(57)摘要

一种用于修正由时域解码器解码的时域激励的合成的方法和设备,其中解码的时域激励的合成被分类到多个类别之一。将解码的时域激励转换成频域激励,并作为解码的时域激励的合成被分类到的类别的函数,对频域激励进行修正。将修正的频域激励转换成修正的时域激励,并向合成滤波器提供修正的时域激励,以产生解码的时域激励的修正的合成。



[接上页]

(51) Int. Cl.

G10L 25/81(2013.01)

G10L 19/22(2013.01)

G10L 25/93(2013.01)

G10L 25/78(2013.01)

1. 一种用于修正由时域解码器解码的时域激励的通用音频合成的设备,包括:  
用于将解码的时域激励转换到频域激励的第一转换器;  
用于频域激励的修正器,其中频域激励被划分成多个频带,每个频带划分成多个频率区间,并且其中所述修正器包括:使用随频带变化的归一化因子对频域激励进行归一化以产生归一化的频域激励的归一化器;  
用于将修正的频域激励转换到修正的时域激励的第二转换器;以及  
合成滤波器,被提供修正的时域激励以产生解码的时域激励的修正的合成。
2. 如权利要求1所述的设备,其中修正器包括对包含多个频率区间的频带中的归一化的频域激励的最大值的一部分以下的频率区间进行归零的归零器。
3. 如权利要求1所述的设备,其中所述修正器包括:  
使用修正前和修正后的频域激励的能量计算每个频带的匹配增益的第一计算器。
4. 如权利要求3所述的设备,其中所述修正器,包括对于第一频率以下的频带:  
将匹配增益应用到归一化的频域激励以产生修正的频域激励的第二计算器。
5. 如权利要求3所述的设备,对于第一较低频率和第二较高频率之间的频带包括:  
用于查找归一化的频域激励在每个频带的最大值的查找器;  
放大器,用于当频率区间中归一化的频域激励等于或高于与频带的所述最大值成比例的值时,在每个频率区间将匹配增益放大一放大因子;以及  
用于将放大的匹配增益应用到频率区间中归一化的频域激励以在所述频率区间中产生修正的频域激励的第二计算器。
6. 如权利要求3所述的设备,对于第一较低频率和第二较高频率之间的频带包括:  
用于查找归一化的频域激励在每个频带的最大值的查找器;  
衰减器,用于当频率区间中归一化的频域激励低于与频带的所述最大值成比例的值时,在每个频带的频率区间将匹配增益衰减一衰减因子;  
用于将衰减的匹配增益应用到所述频率区间中归一化的频域激励以在所述频率区间中产生修正的频域激励的第二计算器。
7. 如权利要求3所述的设备,对于给定频率以上的频带包括:  
用于查找归一化的频域激励在每个频带的最大值的查找器;  
用于当频率区间中归一化的频域激励高于与频带的所述最大值成比例的值时,计算匹配增益的倾斜的第二计算器,所述第二计算器将计算出的倾斜应用到匹配增益;以及  
用于将已经应用了计算出的倾斜的匹配增益应用到所述频率区间中归一化的频域激励以在所述频率区间中产生修正的频域激励的第三计算器。
8. 如权利要求3所述的设备,对于给定频率以上的频带包括:  
用于查找归一化的频域激励在每个频带的最大值的查找器;  
衰减器,用于当频域区间中的归一化的频域激励低于与频带的最大值成比例的值时,在频带的每个频率区间将匹配增益衰减一放大因子;以及  
用于将衰减的匹配增益应用到所述频率区间中归一化的频域激励以在所述频率区间中产生修正的频域激励的第二计算器。
9. 一种用于解码通过编码参数编码的声音信号的设备,包括:  
响应于声音信号编码参数对时域激励进行解码的解码器;

合成滤波器,响应于解码的时域激励产生所述时域激励的合成;以及  
根据权利要求1至8中任一项所述的设备。

10.一种用于修正由时域解码器解码的时域激励的通用音频合成的方法,包括:

将解码的时域激励转换到频域激励;

修正频域激励,其中频域激励被划分成多个频带,每个频带划分成多个频率区间,并且  
其中修正频域激励包括:使用随频带变化的归一化因子归一化频域激励以产生归一化的频  
域激励;

将修正的频域激励转换成修正的时域激励;以及

合成修正的时域激励以产生解码的时域激励的修正的合成。

11.如权利要求10所述的方法,其中修正频域激励包括对包含多个频率区间的频带中的  
归一化的频域激励的最大值的一部分以下的频率区间进行归零。

12.如权利要求10所述的方法,其中修正频域激励包括:

使用修正前和修正后的频域激励的能量计算每个频带的匹配增益。

13.如权利要求12所述的方法,其中修正频域激励包括对于第一频率以下的频带,将匹  
配增益应用到归一化的频域激励以产生修正的频域激励。

14.如权利要求12所述的方法,包括对于第一较低频率和第二较高频率之间的频带:

查找归一化的频域激励在每个频带的最大值;

当频率区间中归一化的频域激励等于或高于与频带的所述最大值成比例的值时,在每  
个频率区间将匹配增益放大一放大因子;以及

将放大的匹配增益应用到频率区间中归一化的频域激励以在所述频率区间中产生修  
正的频域激励。

15.如权利要求12所述的方法,包括对第一较低频率和第二较高频率之间的频带:

查找归一化的频域激励在每个频带的最大值;

当频率区间中归一化的频域激励低于与频带的所述最大值成比例的值时,在频带的每  
个频率区间将匹配增益衰减一衰减因子;

将衰减的匹配增益应用到所述频率区间中归一化的频域激励以在所述频率区间中产  
生修正的频域激励。

16.如权利要求12所述的方法,包括对于给定频率以上的频带:

查找归一化的频域激励在每个频带的最大值;

当频率区间中归一化的频域激励高于与频带的所述最大值成比例的值时,计算匹配增  
益的倾斜,并将计算出的倾斜应用到匹配增益;以及

将已经应用了计算出的倾斜的匹配增益应用到所述频率区间中归一化的频域激励以  
在所述频率区间中产生修正的频域激励。

17.如权利要求12所述的方法,包括对给定频率以上的频带:

查找归一化的频域激励在每个频带的最大值;

当频域区间中的归一化的频域激励低于与频带的最大值成比例的值时,在频带的每个  
频率区间将匹配增益衰减一放大因子;以及

将衰减的匹配增益应用到所述频率区间中归一化的频域激励,以在所述频率区间中产  
生修正的频域激励。

18.一种用于解码通过编码参数编码的声音信号的方法,包括:  
响应于声音信号编码参数解码时域激励;  
合成解码的时域激励以产生所述时域激励的合成;以及  
根据权利要求10至17中任一项所述的方法。

## 用于修正通用音频合成的设备及其方法

[0001] 本申请是申请日为2012年11月1日、申请号为201280065936.1、发明名称为“改善低速率码激励线性预测解码器的非语音内容”的发明专利申请的分案申请。

### 技术领域

[0002] 本公开涉及一种装置和方法，用于修正由时域解码器解码的时域激励的合成。

### 背景技术

[0003] 现有技术的会话编解码器可以用非常好的质量表现具有大约8kbps比特率的干净语音信号和16kbps比特率的近似透明度。为了维持这样的高语音质量，即使在低比特率，也可以使用多模态编码方案。通常，输入声音信号被分离成反映其特性的不同类别。例如，不同的类别可以包括浊音(voiced)、清音(unvoiced)和起始(onset)。编解码器对所有这些类别使用优化的不同的编码模式。

[0004] 然而，一些部署的语音编解码器不使用这种多模态方法，导致对不同于干净语音的声音信号的次优质量(特别是在低比特率)。由于比特流被标准化，并且对比特流的任何修正都将破坏编解码器的互操作性的事实，因此在部署编解码器时，很难对编码器进行修正。然而，对解码器进行修正可以被实现用来改善在接收器侧感知到的质量。

### 发明内容

[0005] 根据第一方面，本公开涉及一种用于修正由时域解码器解码的时域激励的合成的设备，包括：分类器，用于将解码的时域激励的合成分类到多个类别之一；用于将解码的时域激励转换到频域激励的转换器；修正器，作为通过分类器将解码的时域激励的合成分类到的类别的函数来修正频域激励；用于将修正的频域激励转换到修正的时域激励的转换器；以及合成滤波器，被提供修正的时域激励以产生解码的时域激励的修正的合成。

[0006] 根据另一个方面，本公开涉及一种用于解码通过编码参数编码的声音信号的设备，包括：响应于声音信号编码参数对时域激励进行解码的解码器；合成滤波器，响应于解码的时域激励产生所述时域激励的合成；以及上述用于修正时域激励的合成的设备。

[0007] 根据第三方面，本公开涉及一种用于修正由时域解码器解码的时域激励的合成的方法，包括：将解码的时域激励的合成分类到多个类别之一中；将解码的时域激励转换到频域激励；作为解码的时域激励的合成被分类到的类别的函数来修改频域激励；将修正的频域激励转换成修正的时域激励；以及合成修正的时域激励以产生解码的时域激励的修正的合成。

[0008] 根据进一步的方面，本公开涉及一种用于解码通过编码参数编码的声音信号的方法，包括：响应于声音信号编码参数解码时域激励；合成解码的时域激励以产生所述时域激励的合成；以及上述用于修正时域激励的合成的方法。

[0009] 在阅读参照附图以非限制性示例的方式给出的下面的非限制性描述之后，用于修正时域激励的合成的设备和方法的前述和其他特征将变得更加明显。

## 附图说明

[0010] 在附图中：

[0011] 图1是示出用于无效 (inactive) 和有效 (active) 的清音帧改善的CELP解码器的变型的简化示意图；

[0012] 图2是示出用于无效和有效的清音帧改善的CELP解码器变型的详细示意图；

[0013] 图3是示出用于通用音频帧改善的CELP解码器的变型的简化示意图；以及

[0014] 图4是示出用于通用音频帧改善的CELP解码器变型的详细示意图。

## 具体实施方式

[0015] 本公开涉及一种在解码器侧实现多模态解码以维持互操作性并提升感知质量的方法。在本公开中,虽然如在参考文献[3GPP TS 26.190,《Adaptive Multi-Rate-Wideband (AMR-WB) speech codec;Transcoding functions》]中描述的AMR-WB用作示例性示例,但应该切记的是,这种方法也可以应用到其他类型的低比特率语音解码器,该参考文献的全部内容通过引用并入本文。

[0016] 参照图1,为了实现这种多模态解码,时域激励解码器102首先对接收到的比特流101进行整体解码,例如AMR-WB比特流,以获得完整的时域码本激励线性预测 (CELP) 解码的激励。解码的时域激励通过线性预测 (LP) 合成滤波器103进行处理,以获得解码器内部采样频率的语音/声音信号时域合成。对于AMR-WB,该内部采样频率为12.8kHz,但对另一编解码器,它可能是不同的。

[0017] 来自LP合成滤波器103的当前帧的时域合成通过分类器104-105-106-301 (图1、图2和图3) 进行处理,其中向所述分类器提供来自比特流101的话音有效性检测 (VAD) 信息109。分类器104-105-106-301分析时域合成并将其分类成无效语音、有效浊音语音、有效清音或通用音频。无效语音 (在1051检测到) 包括语音突发、表示主动语音突发期间具有浊音特性的帧的有效浊音语音 (在1061检测到)、表示在语音突发期间具有清音特性的帧的有效清音语音 (在1062检测到)、以及表示音乐或混响语音的通用音频 (在3010检测到) 之间的所有的背景噪音。其他的类别可以添加或从上述类别得到。所公开的方法的目的是改进尤其 (但并非排他性地) 是无效的语音、有效的清音语音和通用音频的感知质量。

[0018] 一旦确定了时域合成的类别,转换器/修正器107就使用非重叠频率变换将来自时域激励解码器102的解码的激励转换到频域。也可以使用重叠变换,但它意味着增加了端到端延迟,这在大多数情况下是不希望的。然后,在转换器/修正器107中,将激励的频率表示分离到不同的频带。频带可以有固定的大小,可以依赖于临界频带[[J.D. Johnston, “Transform coding of audio signal using perceptual noise criteria”, IEEE J. Select. Areas Commun., 第6卷, 第314-323页, 1988年2月] (其全部内容通过引入并入本文), 或任何其他组合。然后,计算每个频带的能量,并将其保存在转换器/修正器107的存储器中,以在整形 (resample) 处理后使用,以确保修正不改变全局的帧能量级。

[0019] 转换器/修正器107执行的频域中的激励的修正可能会随着合成的分类而不同。对于无效的语音和有效的清音语音,整形可以包括对低频归一化再添加噪声,以及仅用噪声替换高频内容。解码的时域合成的截止频率、低频和低频之间的高频界限可以固定在约1至1.2kHz的值。当在未修正的帧和修正的帧之间进行切换时,保持解码的时域合成的一些低

频内容以防止伪像。也可以通过选择频率区间 (frequency bin) 作为来自时域激励解码器 102 的解码的基音 (pitch) 的函数而使得截止频率逐帧可变化。修正处理具有除去与低比特率语音编解码器相关联的电噪声种类的效果。修正处理后,应用每频带的增益匹配,以取回每个频带的初始能量级,其中稍微增加 6kHz 以上的频率的能量,以补偿在这些频率上的 LP 滤波器增益下降。

[0020] 对于被归类为通用音频的帧,转换器/修正器 107 中的处理是不同的。首先,对所有频带的每一个频带进行归一化。在归一化操作中,频带中低于频带内最大频率值的一部分的所有区间都被设置为零。对于更高的频段,每个频段内更多的区间归零。这以较高的比特预算模拟了频率量化方案,但更多的比特分配给了较低的频率。归一化处理后,可以应用噪声填充,用随机噪声替换归零的区间,但取决于不同的比特率,并不总是使用噪声填充。修正处理后,应用每频带的增益匹配,以取回每个频带的初始能量级,并沿频带应用取决于比特率的倾斜校正,以补偿在通用音频输入情形中对 LP 滤波器的系统性低估。通用音频路径的另一个不同来自于增益匹配并非应用在所有的频率区间上的事实。由于通用音频的频谱通常比语音更加多尖峰,因此当可以识别频谱脉冲并对其加重时,改善了感知质量。要做到这一点,只对频带内的最高能量区间应用具有倾斜校正的全增益匹配。对于最低能量区间,对这些区间只应用一部分增益匹配。这导致增加频谱动态。

[0021] 在激励频率整形和增益匹配之后,转换器/修正器 107 应用逆频率变换,以获得修正的时域激励。该修正的激励通过 LP 合成滤波器 108 处理,以得到修正的时域合成。在去加重滤波器和重采样器 112 中最终去加重和重采样至 16kHz (对于 AMR-WB 的示例) 之前,取决于时域解码合成的分类,覆写器 110 简单地用来自 LP 合成滤波器 108 的修正后的时域合成覆写来自 LP 合成滤波器 103 的时域解码合成。

[0022] 在无效的语音的情形中,与有效的清音语音的修正相比,唯一的区别是使用平滑器 111 用于平滑 LP 合成滤波器 108 以给出更平滑的噪声变化。其余的修正都与有效的清音路径相同。在下面的文本中,参照图 2 描述实现公开的方法的更详细的示例。

[0023] 1) 信号分类

[0024] 参照图 2,分类器 104-105-106-301 对于应用了修正的比特率在解码器执行如上文描述的语音/声音信号的时域合成的分类 1021。为了简化附图的目的,图 2 没有示出 LP 滤波器 103。解码器处的分类与如在参考文献 [Milan Jelinek 和 Philippe Gournay; PCT 专利申请 W003102921A1, “A method and device for efficient frame erasure concealment in linear predictive based speech codecs”] 和 [T. Vaillancourt 等人, PCT 专利申请 W02007073604A1, “Method and device for efficient frame erasure concealment in speech codecs”] 中描述的相似 (这些参考文献的全部内容通过引用并入本文), 加上对通用音频检测的一些适配。下面的参数用于在解码器处对帧进行分类: 归一化的相关性  $r_x$ 、频谱倾斜度量  $e_t$ 、基音稳定性计数器  $p_c$ 、在当前帧  $E_s$  结束处声音信号的相对帧能量和过零计数器  $z_c$ 。这些用于对信号进行分类的参数的计算在下面说明。

[0025] 归一化的相关性  $r_x$  在帧结束处基于语音/声音信号的时域合成  $S_{out}(n)$  计算。使用来自时域激励解码器 102 的最后一个子帧的基音滞后 (lag)。更具体地,归一化的相关性  $r_x$  基音同步地计算如下:



$$[0026] \quad r_x = \frac{\sum_{i=0}^{T-1} x(t+i)x(t+i-T)}{\sqrt{\sum_{i=0}^{T-1} x^2(t+i) \sum_{i=0}^{T-1} x^2(t+i-T)}} \quad (1)$$

[0027] 其中  $x(n) = S_{out}(n)$ ,  $T$  是最后一个子帧的基音滞后,  $t = L - T$ ,  $L$  是帧大小。如果最后一个子帧的基音滞后大于  $3N/2$  ( $N$  是子帧的大小), 那么将  $T$  设置为最后两个子帧的平均基音滞后。

[0028] 因此, 归一化的相关性  $r_x$  是使用语音/声音信号时域合成  $S_{out}(n)$  计算得到的。对于低于子帧大小 (64 个采样) 的基音滞后, 归一化的相关性在时刻  $t = L - T$  和  $t = L - 2T$  计算两次, 并且归一化的相关性  $r_x$  作为这两次计算的平均值给出。

[0029] 频谱倾斜参数  $e_t$  包含有关能量的频率分布的信息。作为非限定性的示例, 解码器处的频谱倾斜被估计为时域合成的第一归一化自相关系数。它基于最后的 3 个子帧计算为:

$$[0030] \quad e_t = \frac{\sum_{i=N}^{L-1} x(i)x(i-1)}{\sum_{i=N}^{L-1} x^2(i)} \quad (2)$$

[0031] 其中  $x(n) = S_{out}(n)$  是时域合成信号,  $N$  是子帧大小,  $L$  是帧大小 (在 AMR-WB 的示例中,  $N=64$ ,  $L=256$ )。

[0032] 基音稳定性计数器  $pc$  对基音周期的变化进行评估。它在解码器处如下计算:  $pc = |p_3 + p_2 - p_1 - p_0|$  (3)

[0033] 值  $P_0, P_1, P_2$  和  $P_3$  对应于来自当前帧的 4 个子帧的闭环基音滞后 (在 AMR-WB 的示例中)。

[0034] 相对帧能量  $E_s$  计算为以 dB 为单位的当前帧能量  $E_f$  和其长期平均值  $E_{lt}$  之间的差

$$[0035] \quad E_s = E_f - E_{lt} \quad (4)$$

[0036] 其中当前帧能量  $E_f$  是在帧结束处基音同步地计算得到的以 dB 为单位的时域合成  $S_{out}(n)$  的能量

$$[0037] \quad E_f = 10 \log_{10} \left( \frac{1}{T} \sum_{i=0}^{T-1} s_{out}^2(i + L - T) \right) \quad (5)$$

[0038] 其中  $L=256$  (在 AMR-WB 的示例中) 是帧的长度,  $T$  是最后两个子帧的平均基音滞后。如果  $T$  小于子帧的大小, 那么将  $T$  设置为  $2T$  (对于短基音滞后使用两个基音周期计算得到的能量)。

[0039] 使用下面的关系式更新有效语音帧的长期平均能量:

$$[0040] \quad E_{ff} = 0.99E_{ff} + 0.01E_f \quad (6)$$

[0041] 最后一个参数是在时域合成  $S_{out}(n)$  的一个帧上计算得到的过零计数器  $zc$ 。作为非限制性的示例, 过零计数器  $zc$  计数在该间隔期间时域合成的符号从正变为负的次数。

[0042] 为了使分类更鲁棒, 将分类参数放在一起考虑形成评价 (merit) 函数  $f_m$ 。为了这个目的, 首先使用线性函数对分类参数进行缩放。让我们考虑参数  $P_x$ , 其缩放版本使用下式获得:

$$[0043] \quad p^s = k_p \cdot p_x + c_p \quad (7)$$

[0044] 缩放后的基音稳定性计数器  $pc$  被截短 (clip) 在 0 和 1 之间。对于每个参数已经通过

实验找到函数系数 $K_p$ 和 $C_p$ 。在该实现的示例中使用的值总结在表1中:

[0045] 表1:解码器处的帧分类参数和其相应的缩放函数的系数

	参数	含义	$K_p$	$C_p$
[0046]	$r_x$	归一化的相关性	0.8547	0.2479
	$e_t$	频谱倾斜	0.8333	0.2917
[0047]	$pc$	基音稳定性计数器	-0.0357	1.6074
	$E_s$	相对帧能量	0.04	0.56
	$zc$	过零计数器	-0.04	2.52

[0048] 评价函数定义为:  $f_m = \frac{1}{6}(2 \cdot r_x^s + e_t^s + pc^s + E_s^s + zc^s)$  (8) 其中上标s表示参数的缩放版本。

[0049] 然后,使用评价函数 $f_m$ 并遵循在表2中总结的规则进行帧的分类:

[0050] 表2:解码器处的信号分类规则

	前一帧分类	规则	当前帧分类
[0051]	起始 浊音 浊音过渡 (voiced transition) 人造起始 (artificial onset) 通用音频声音	$f_m \geq 0.63$	浊音
		$0.39 \leq f_m < 0.63$	浊音过渡
		$f_m < 0.39$	清音
	清音过渡 (unvoiced transition) 清音	$f_m > 0.56$	起始
		$0.56 \geq f_m > 0.45$	清音过渡
		$f_m \leq 0.45$	清音
	当前帧 VAD 信息		
		$VAD = 0$	清音

[0052] 除此分类之外,与编码器的话音有效性检测(VAD)相关的信息109可被发送至比特流101中(图1),与在AMR-WB的示例中的情形一样。因此,一比特被发送到比特流101以指定编码器是否将当前帧视为有效内容( $VAD=1$ )或无效内容(背景噪声, $VAD=0$ )。当VAD信息表明内容是无效的时,分类器部分104,105,106和301然后将分类覆写为清音。

[0053] 分类方案还包括通用的音频检测(见分类器部分301,图3)。通用音频类别包括音乐、混响语音,还可以包括背景音乐。分类的第二个步骤允许分类器104-105-106-301以较好的置信度确定当前帧可以被归类为通用音频。使用两个参数来实现该第二分类步骤。一

个参数是如等式 (5) 中表达的总帧能量 $E_f$ 。

[0054] 首先,过去四十 (40) 个总帧能量变化的平均值 $\overline{E}_{df}$ 用下面的关系式计算得到:

$$[0055] \quad \overline{E}_{df} = \frac{\sum_{t=-40}^{t=-1} \Delta_E^t}{40}; \text{ 其中 } \Delta_E^t = E_f^t - E_f^{(t-1)} \quad (9)$$

[0056] 然后,最后十五 (15) 个帧上的能量变化历史的统计偏差 $\sigma_E$ 用下面的关系式确定:

$$[0057] \quad \sigma_E = 0.7745967 \cdot \sqrt{\sum_{t=-15}^{t=-1} \frac{(\Delta_E^t - \overline{E}_{df})^2}{15}} \quad (10)$$

[0058] 得到的偏差 $\sigma_E$ 给出对解码的合成的能量稳定性的指示。通常情况下,音乐比语音具有更高的能量稳定性(较低的能量变化历史的统计偏差)。

[0059] 此外,第一个步骤的分类被用来评估当如在等式 (5) 中表达的帧能量 $E_f$ 大于-12dB 时被分类为清音 $N_{uv}$ 的两个帧之间的间隔。当帧被分类为清音,并且帧能量 $E_f$ 大于-9dB时,意味着信号是清音而不是无声(silent),如果如在等式 (6) 中表达的长期有效语音能量 $E_{lt}$ 小于40dB,那么将清音间隔计数器设置为16,否则将清音间隔计数器 $N_{uv}$ 减小8。对于有效语音信号,计数器 $N_{uv}$ 也被限制在0和300之间,对于无效语音信号,被限制在0和125之间。应该提醒注意的是,在说明性的示例中,有效和无效语音信号之间的差可以从包含在比特流101中的话音有效性检测VAD信息推导得到。

[0060] 对于有效语音信号,从该清音帧计数器得到长期平均值:

$$[0061] \quad N_{uv_{lt}} = 0.9 \cdot N_{uv_{lt}} + 0.1 \cdot N_{uv} \quad (11)$$

[0062] 对于无效的语音信号则如下:

$$[0063] \quad N_{uv_{lt}} = 0.95 \cdot N_{uv_{lt}} \quad (12)$$

[0064] 此外,当长期平均值很高,并且偏差 $\sigma_E$ 也高时,例如在实现的当前示例中,当 $N_{uv_{lt}} > 140$ 并且 $\sigma_E > 5$ 时,长期平均值修正如下:

$$[0065] \quad N_{uv_{lt}} = 0.2 \cdot N_{uv_{lt}} + 80 \quad (13)$$

[0066] 分类器104-105-106-301使用有关被分类为清音的帧之间的帧的数量的长期平均值的参数,来确定所述帧是否应被视为通用音频。越多清音帧在时间上接近,帧越可能具有语音特征(更小可能是通用音频)。在说明性的示例中,用来决定帧是否被视为通用音频 $G_A$ 的阈值定义如下:

$$[0067] \quad \text{帧是 } G_A, \text{ 如果: } N_{uv_{lt}} > 140 \text{ 并且 } \Delta_E^t < 12 \quad (14)$$

[0068] 加上在等式 (9) 中定义的参数 $\Delta_E^t$ ,以便不将大的能量变化分类为通用音频,而是将它保持为有效语音。

[0069] 对激励的修正的执行取决于帧的分类,并且对一些类型的帧,根本没有修正。下表3总结了可以进行或不可以进行修正的情况。

[0070] 表3:激励修正的信号类别

[0071]	帧分类	语音有效性检测? 是/否	类别	修正 是/否
	起始 浊音 清音过渡 人造起始	是 (VAD=1)	有效浊音	否
	通用音频声音	是	通用音频	是*
	浊音过渡 清音	是	有效的清音	是
	起始 浊音 清音过渡 人造起始 通用音频声音 浊音过渡 清音	否	无效音频	是

[0072] \*可以取决于对通用音频类别进行或不进行修正。例如,可以仅当通用音频是无效的时候对其进行修正,或仅当通用音频是有效的时候进行修正,始终进行修正或根本不进行修正。

[0073] 2) 频率变换

[0074] 在频域修正阶段,需要将激励表示到变换域。例如,通过使用给出25Hz频率分辨率的II型DCT(离散余弦变换)的转换器/修正器107的时域到频域转换器201实现时间到频率转换,但也可以使用任何其他合适的变换。如果使用另一种变换,那么频率分辨率(上面定义的)、频带的数目和每个频带的频率区间的数目(下文进一步定义)可能需要相应地修改。在时域到频域转换器201中计算的时域CELP激励 $f_e$ 的频率表示如下给出:

$$[0075] \quad f_e(k) = \begin{cases} \sqrt{\frac{1}{L}} \cdot \sum_{n=0}^{L-1} e_{td}(n), & k = 0 \\ \sqrt{\frac{2}{L}} \cdot \sum_{n=0}^{L-1} e_{td}(n) \cdot \cos\left(\frac{\pi}{L}\left(n + \frac{1}{2}\right)k\right), & 1 \leq k \leq L-1 \end{cases} \quad (15)$$

[0076] 其中 $e_{td}(n)$ 是时域CELP激励, $L$ 是帧长度。在AMR-WB的示例中,对于对应的内采样频率12.8kHz帧长度是256个采样。

[0077] 在时域CELP解码器,诸如102中,时域激励信号由以下给出:

$$[0078] \quad e_{td}(n) = bv(n) + gc(n) \quad (15)$$

[0079] 其中 $V(n)$ 是自适应码本贡献, $b$ 是自适应码本增益, $c(n)$ 是固定码本贡献, $g$ 是固定码本增益。

[0080] 3) 每个频带的能量分析

[0081] 在对时域激励进行任何修正之前,转换器/修正器107包括增益计算器208-209-210,其本身包括子计算器209,用来计算频域激励的每个频带的能量 $E_b$ 并将计算得到的每个频带的能量 $E_b$ 保存在存储器中以用于激励频谱整形后的能量调整。对于12.8kHz的采样频率,可以通过子计算器209如下计算能量:

$$E_b(i) = \sqrt{\sum_{j=C_{Bb}(i)}^{j=C_{Bb}(i)+B_b(i)} f_e(j)^2} \quad (16)$$

其中 $C_{Bb}$ 代表每个频带的累积频率区间, $B_b$ 为每个频带的区间数量,定义为:

[0082]  $B_b = \{4, 4, 4, 4, 4, 5, 6, 6, 6, 8, 8, 10, 11, 13, 15, 18, 22, 16, 16, 20, 20 \cdot 20, 16\}$

[0083]  $C_{Bb} = \left\{ \begin{array}{l} 0, 8, 12, 16, 20, 25, 31, 37, 43, 51, 59, 69, 80, 93, \\ 108, 126, 148, 164, 180, 200, 220, 240 \end{array} \right\}$

[0084] 低频带可以对应于临界(critical)音频频带,如在Milan Jelinek和Philippe Gournay的PCT专利申请W003102921A1,“A method and device for efficient frame erasure concealment in linear predictive based speech codecs”中描述的,其全部内容通过引用并入本文,但3700Hz以上的频带可以短一点,以更好地匹配这些频带中可能的频谱能量变化。频谱频带的任何其他配置也是可能的。

[0085] 4) 无效和有效的清音帧的激励修正

[0086] a) 时域贡献的截止频率对比噪声填充

[0087] 为了实现无效帧和有效的清音帧的未修正激励和修正后激励之间的透明切换,至少保持时域激励贡献的较低频率。转换器/修正器107包括截止频率计算器203,用来确定停止使用时域贡献的频率,截止频率 $f_c$ 具有最小值1.2kHz。这意味着,始终保持解码的激励的第一个1.2kHz,并且取决于来自时域激励解码器102的解码的基音值,截止频率可以更高。8次谐波从所有子帧的最低基音计算得到,时域贡献保持高达该8次谐波。对8次谐波的估计计算如下:

$$h_{8th} = \frac{(8 \cdot F_s)}{\min_{0 \leq i < N_{sub}}(T(i))} \quad (17)$$

[0089] 其中 $F_s = 12800\text{Hz}$ ,  $N_{sub}$ 是子帧的数量, $T$ 是解码的子帧基音。对于所有的 $i < N_b$ ,其中 $N_b$ 是被包括在频率范围 $L_f$ 中的最大频带,通过搜索最高的频带以找到8次谐波所在的频带来进行验证,对于该最高的频带,依然验证下面的不等式:

$$(h_{8th} \geq L_f(i)) \quad (18)$$

[0091] 其中 $L_f$ 定义为:

$$L_f = \left\{ \begin{array}{l} 175, 275, 375, 475, 600, 750, 900, 1050, 1250, 1450, 1700, 1975, \\ 2300, 2675, 3125, 3675, 4075, 4475, 4975, 5475, 5975, 6375 \end{array} \right\}$$

[0093]  $L_f$ 中的频带的索引将被称为 $i_{gth}$ ,它表明8次谐波可能位于的频段。截止频率计算器203使用下面的关系将最终的截止频率 $f_{tc}$ 计算为1.2kHz和8次谐波可能位于的频带中的最后频率 $(L_f(i_{gth}))$ 之间的较高的频率,:

$$f_{tc} = \max(L_f(i_{gth}), 1.2 \text{ kHz}) \quad (19)$$

[0095] b) 归一化和噪声填充

[0096] 转换器/修正器107还包括将高于截止频率 $f_c$ 的频带的频率区间归零的归零器204。

[0097] 对于无效帧和有效的清音帧,转换器/修正器107的归一化器205使用下面的关系在 $[0,4]$ 之间归一化低于时域CELP激励 $f_e$ 的频率表示的频带的 $f_c$ 的频率区间:

$$[0098] \quad f_{eN}(j) = \begin{cases} \frac{4 \cdot f_e(j)}{\max_{0 \leq i < f_c} (|f_e(i)|)}, & \text{对于 } 0 \leq j < f_c \\ 0, & \text{对于 } f_c \leq j < 256 \end{cases} \quad (20)$$

[0099] 然后,转换器/修正器107包括随机噪声产生器206,用来产生随机噪声,并通过加法器207来进行简单的噪声填充,以在所有的频率区间上以恒定的等级添加噪声。描述噪声添加的函数定义如下:

$$[0100] \quad \text{对于 } j = 0:L-1 \quad f'_{eN}(j) = f_{eN}(j) + 0.75 \cdot \text{rand}(\quad) \quad (21)$$

[0101] 其中 $\text{rand}$ 是限制在-1到1之间的随机数产生器。

[0102] c) 修正后激励频谱的每个频带的能量分析

[0103] 增益计算器208-209-210的子计算器208用与在上面的第3小节描述的相同方法确定在频谱整形后每个频带的能量 $E_b'$ 。

[0104] d) 能量匹配

[0105] 对于无效帧和有效的清音帧,能量匹配仅在于在对其初始值进行激励频谱修正后调整每个频带的能量。对于每个频带 $i$ ,增益计算器208-209-210的子计算器210确定用来应用到频带中的所有区间的匹配增益 $G_b$ ,以如下匹配能量:

$$[0106] \quad G_b(i) = \frac{E_b(i)}{E_b'(i)} \quad (22)$$

[0107] 其中 $E_b(i)$ 是如在子计算器209中使用上面第3小节的方法确定的激励频谱修正之前的每个频带的能量, $E_b'(i)$ 是如在子计算器208中计算的激励频谱修正之后的每个频带的能量。对于特定的频带 $i$ ,如在子计算器210中确定的修正后的(去归一化的)频域激励 $f'_{edN}$ 可以写为:

[0108] 对于 $C_{Bb}(i) \leq j < C_{Bb}(i) + B_b(i)$

$$[0109] \quad f'_{edN}(j) = G_b(i) \cdot f'_{eN}(j) \quad (23)$$

[0110] 其中 $C_{Bb}$ 和 $B_b$ 在上面的第3小节中定义。

[0111] 5) 通用音频帧的激励修正

[0112] a) 归一化和噪声填充

[0113] 现在将参考图3。对于如通过分类部分301确定的通用音频帧,归一化是略有不同的,并通过归一化器302执行。首先,归一化因子 $N_f$ 逐频带变化,对于低频带使用较大的值,对于高频带使用较小的值。这个想法允许在脉冲定位更准确的低频带的较高幅度,以及在脉冲定位不那么准确的较高频带的较低幅度。在这个示例性示例中,根据频带变化的归一化因子 $N_f$ 定义为:

$$[0114] \quad N_f = \{16, 16, 16, 16, 16, 16, 16, 16, 12, 12, 12, 12, 8, 8, 8, 8, 8, 4, 4, 2, 2, 1, 1, 1\}$$

[0115] 对于特定的频带 $i$ ,通用音频帧的时域激励(频域激励) $f_e$ 的频率表示的归一化可

以描述如下：

$$[0116] \quad f_{eN}(j) = \frac{N_f(i) \cdot f_e(j)}{\max_{k=C_{Bb}(i)}^{C_{Bb}(i)+B_b(i)} (|f_e(k)|)} \quad \text{对于 } C_{Bb}(i) \leq j < C_{Bb}(i) + B_b(i) \quad (24)$$

[0117] 其中 $B_b$ 是每个频带的区间的数量,每个频带的累积频率区间是 $C_{Bb}$ , $f_{eN}(j)$ 是归一化频域激励。 $B_b$ 和 $C_{Bb}$ 在上面的第3小节中描述。

[0118] 此外,归一化器302包括归零器(未显示),用来对每个频带中 $f_{eN}(j)$ 的最大值的一部分 $Z_f$ 以下的所有频率区间进行归零以得到 $f'_{eN}(j)$ ：

$$[0119] \quad f'_{eN}(j) = \begin{cases} 0 & \text{如果 } (f_{eN}(j) < Z_f(i)) \\ f_{eN}(j) & \text{否则} \end{cases} \quad \text{对于 } C_{Bb}(i) \leq j < C_{Bb}(i) + B_b(i) \quad (25)$$

[0120] 其中 $Z_f$ 可以表示为：

$$[0121] \quad Z_f = \{1, 0.5, 0.5, 0.5\}$$

[0122] 如果想要提高频谱的多峰性(peakyness),更积极的归零可以通过增大向量 $Z_f$ 的值来执行。

[0123] b) 修正后激励频谱的每个频带的能量分析

[0124] 增益计算器303-304-306的计算器部分303确定用与在上面的第3小节描述的相同方法的频谱整形之后的每频带的能量 $E_b'$ 。

[0125] c) 能量匹配

[0126] 图3示出了增益计算器303-304-306,图4更详细描述了该增益计算器的计算器部分306。

[0127] 对于通用音频帧,能量匹配更复杂,因为它的目的也是提高频谱动态。对于每个频带 $i$ ,增益计算器303-304-306的计算器部分306的子计算器413计算如在等式(22)中类似地定义的估计的增益 $G_e$ ：

$$[0128] \quad G_e(i) = \frac{E_b(i)}{E_b'(i)} \quad (26)$$

[0129] 其中 $E_b(i)$ 是如在计算器部分304中使用如上面第3小节描述的方法确定的激励频谱修正之前的每个频带的能量, $E_b'(i)$ 是如在计算器部分303中计算的激励频谱修正之后的每个频带的能量。

[0130] 计算器部分306的子计算器414使用下面的关系将增益 $G$ 应用到来自归一化器302和频谱分割器401-420的归一化频域激励 $f'_{eN}$ 的第一个400Hz(或最先的4个频带)以提供修正的(去归一化的)频域激励 $f'_{edN}$ ：

$$[0131] \quad f'_{edN}(j) = G_e(i) \cdot f'_{eN}(j), \text{ 对于 } C_{Bb}(i) \leq j < C_{Bb}(i) + B_b(i) \quad 1 \leq i < 4 \quad (27)$$

[0132] 查找器404确定在400Hz以上的每个频带 $j$ 的最大值 $\max_{a \leq j < b} (|f_{eN}(j)|)$ ,其中 $a = C_{Bb}(i)$ , $b = C_{Bb}(i) + B_b(i)$ 在上面的第3小节中定义。

[0133] 对于包含在归一化频域激励的400Hz和2kHz(频带4至12)之间的频带(参见模块420和450),如果归一化频域激励在频率区间 $f'_{eN}(j) \geq 0.86 \max_{a \leq j < b} (|f_{eN}(j)|)$ 中(参见模块451),则放大器402如在等式(28)的上面一行所示将来自子计算器413的增益 $G_e$ 放大因子

1.1。子计算器403根据等式(28)的第一行将来自放大器402的放大的增益应用到频率区间中的归一化频谱激励 $f'_{eN}$ 以获得修正的(去归一化的)频域激励 $f'_{edN}$ 。

[0134] 再次,对于包含在归一化频域激励的400Hz和2kHz(频带4至12)之间的频带(参见模块420和450),如果归一化频域激励在频率区间 $f'_{eN}(j) < 0.86 \max_{a \leq j < b} (|f_{eN}(j)|)$ 中(参见模块451),则衰减器405如在等式(28)的下面一行所示将来自子计算器413的增益 $G_e$ 衰减因子0.86。子计算器406根据等式(28)的下面一行将来自衰减器405的衰减的增益应用到频率区间中的归一化频谱激励 $f'_{eN}$ 以获得修正的(去归一化的)频域激励 $f'_{edN}$ 。

[0135] 用于总结,修正后的(去归一化的)频谱激励 $f'_{edN}$ 如下给出:

$$[0136] \quad f'_{edN}(j) = \begin{cases} 1.1 \cdot G_e(i) \cdot f'_{eN}(j), & \text{如果 } f'_{eN}(j) \geq 0.86 \cdot \max_{a \leq j < b} (|f_{eN}(j)|) \\ 0.86 \cdot G_e(i) \cdot f'_{eN}(j), & \text{如果 } f'_{eN}(j) < 0.86 \cdot \max_{a \leq j < b} (|f_{eN}(j)|) \end{cases} \quad (28)$$

[0137] 最后,对于频谱中更高的部分,在该示例中为归一化频域激励2kHz以上的频带(频带 $>12$ ) (参见模块420和450),如果归一化频域激励在频率区间 $f'_{eN}(j) \geq 0.86 \max_{a \leq j < b} (|f_{eN}(j)|)$ 中(参见模块452),那么将作为频带 $i$ 的函数并且也可以是比特率的函数的倾斜添加到增益 $G_e$ ,以补偿LPC滤波器过低的能量估计。每个频带的倾斜值 $\delta(i)$ 表示为:

$$[0138] \quad \delta(i) = 1.5 \cdot G_e(i) \cdot \frac{(j-12)}{32} \quad (29)$$

[0139] 倾斜通过倾斜计算器407-408计算得到,并通过子计算器409根据等式(30)的上面一行按频率区间应用到归一化频域激励 $f'_{eN}$ ,以得到修正的(去归一化的)频域激励 $f'_{edN}$ 。

[0140] 再次,对于频谱中更高的部分,在该示例性示例中为归一化频域激励2kHz以上的频带(频带 $>12$ ) (参见模块420和450),如果归一化频域激励在频率区间 $f'_{eN}(j) < 0.86 \max_{a \leq j < b} (|f_{eN}(j)|)$ 中(参见模块452),那么衰减器410计算由子计算器406根据等式(30)的下面一行按频率区间应用到归一化频谱激励 $f'_{eN}$ 以获得修正的(去归一化的)频域激励 $f'_{edN}$ 的衰减增益 $[f'_{eN}(j) / \max_{a \leq j < b} (|f_{eN}(j)|)]^2$ 。

[0141] 用于总结,去归一化的频谱激励 $f'_{edN}$ 如下确定:

$$[0142] \quad f'_{edN}(j) = \begin{cases} \delta(i) \cdot f'_{eN}(j), & \text{如果 } f'_{eN}(j) \geq 0.86 \cdot \max_{a \leq j < b} (|f_{eN}(j)|) \\ \left( \frac{f'_{eN}(j)}{\max_{a \leq j < b} (|f_{eN}(j)|)} \right)^2 \cdot f'_{eN}(j), & \text{否则} \end{cases} \quad (30)$$

[0143] 其中 $a$ 和 $b$ 在本文的上面有描述。也可以进一步增大应用到最新频带的增益,其中LPC的能量匹配是最差的。

[0144] 6) 逆频率变换

[0145] 合成器453合成来自子计算器414,403,406,409和411的对修正的(去归一化的)频域激励 $f'_{edN}$ 的贡献,以形成完整的修正的(去归一化的)频域激励 $f'_{edN}$ 。

[0146] 在频域处理完成后,将逆频率时间变换202应用于来自合成器453的修正的(去归一化的)频域激励 $f'_{edN}$ ,以找到时域修正后的激励。在这个示例性实施例中,频率到时间转换使用与对于时间到频率转换使用的给出25Hz分辨率的相同的II型DCT的逆过程实现。再次,可以使用任何其他的变换。获得如下的修正的时域激励 $e'_{td}$ :



$$[0147] \quad e'_{id}(k) = \begin{cases} \sqrt{\frac{1}{L}} \cdot \sum_{n=0}^{L-1} f'_{edN}(n), & k = 0 \\ \sqrt{\frac{2}{L}} \cdot \sum_{n=0}^{L-1} f'_{edN}(n) \cdot \cos\left(\frac{\pi}{L}\left(n + \frac{1}{2}\right)k\right), & 1 \leq k \leq L-1 \end{cases} \quad (31)$$

[0148] 其中 $f'_{edN}$ 是修正后的激励的频率表示, $L$ 是帧长度。在这个示例性示例中,对于对应的12.8kHz的内部采样频率,帧长度为256个采样(AMR-WB)。

[0149] 7) 合成滤波和覆写当前的CELP合成

[0150] 一旦完成激励修正,就通过合成滤波器108处理修正后的激励,以获得当前帧的修正的合成。覆写器110使用此修正的合成覆写解码的合成,从而提高感知质量。

[0151] 最后的去加重和重采样到16kHz可以随后在去加重滤波器和重采样器112中进行。



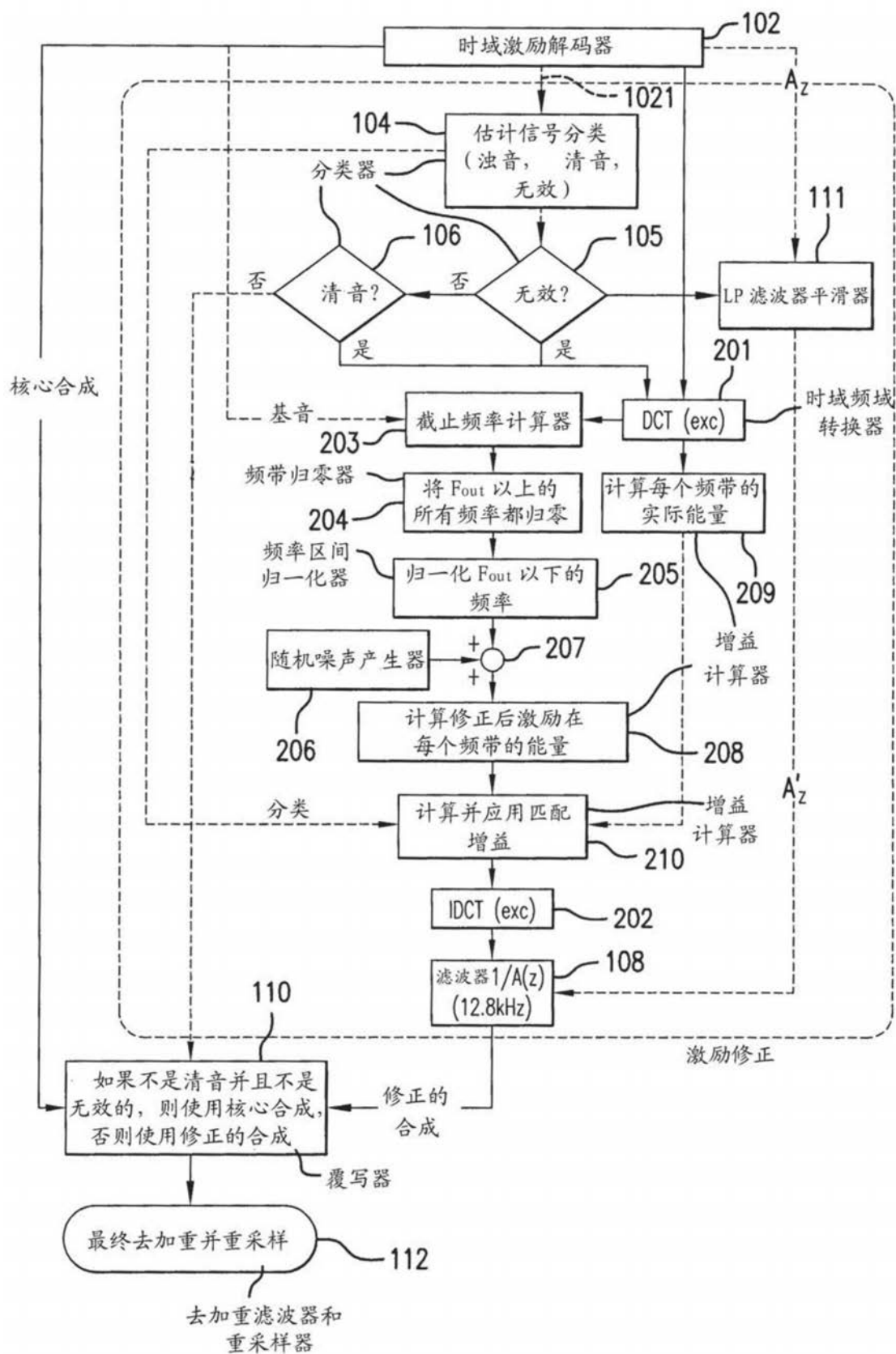


图2

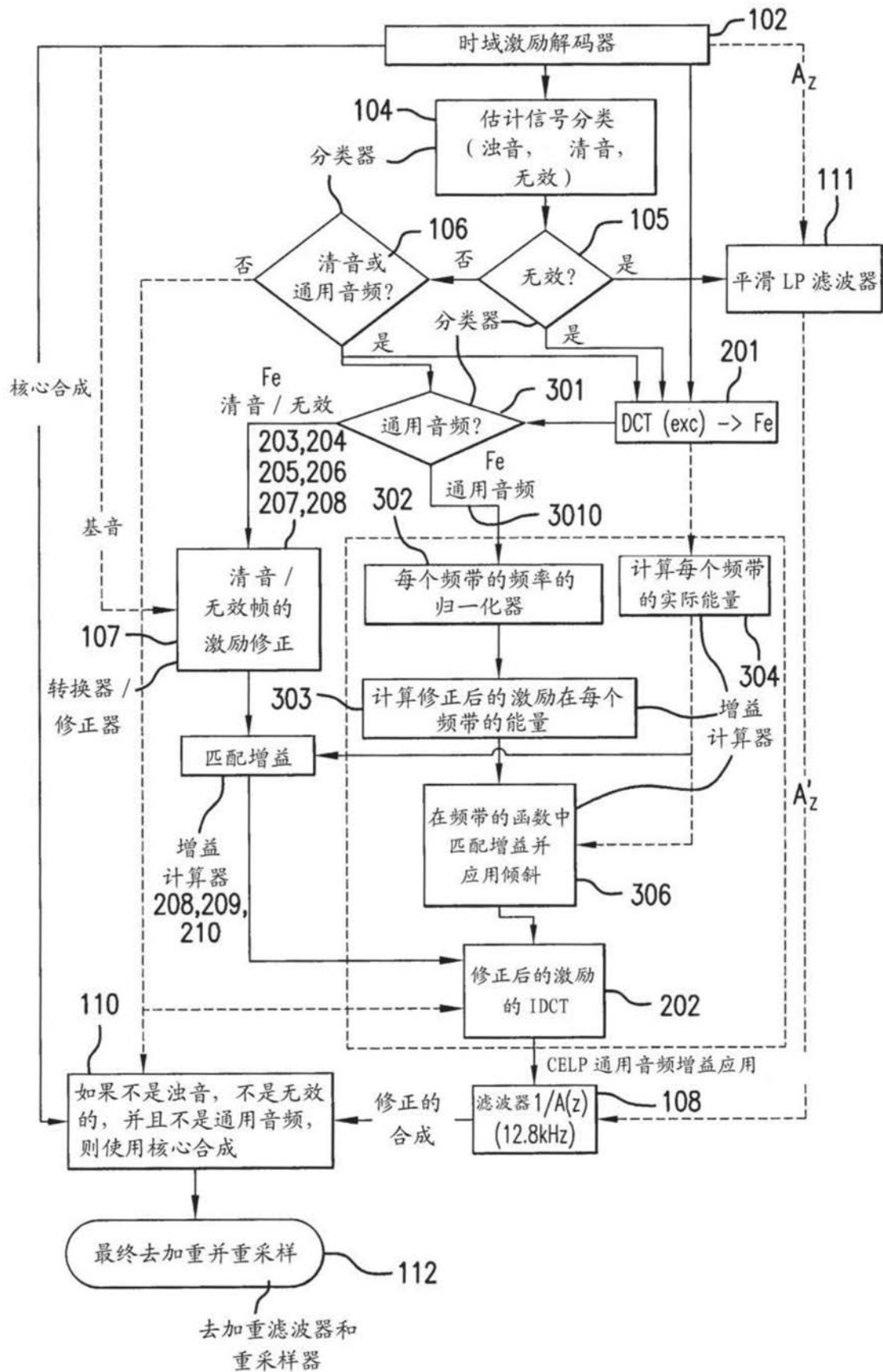


图3

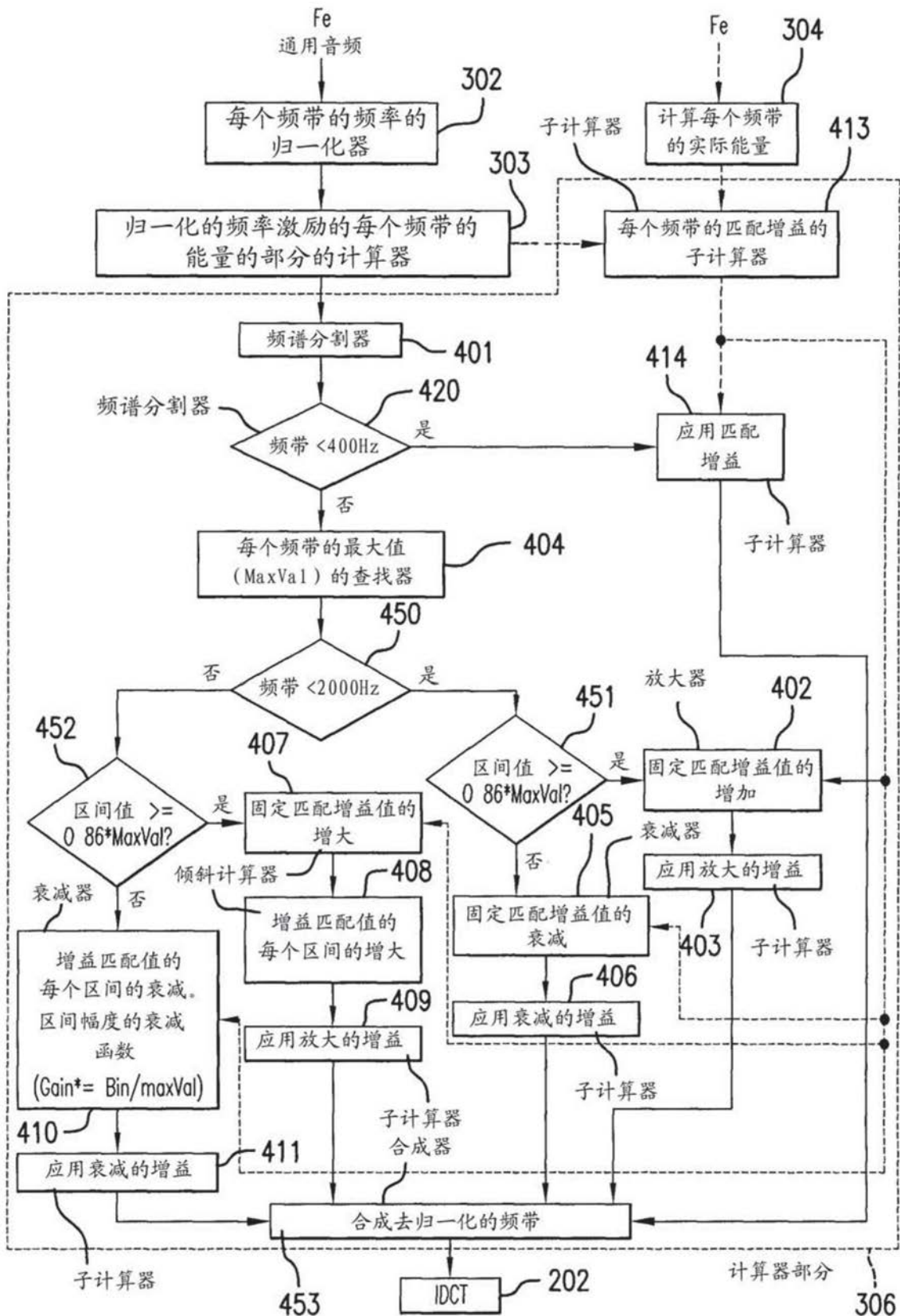


图4