(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 112927725 B
(45)授权公告日 2025.02.18

(21)申请号 202110082923.6
(22)申请日 2015.07.01
(65)同一申请的已公布的文献号
申请公布号 CN 112927725 A
(43)申请公布日 2021.06.08
(30)优先权数据
62/030,121 2014.07.29 US
(62)分案原申请数据
201580040591.8 2015.07.01
(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 马丁·绍尔斯戴德
(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 冯薇
(51)Int.Cl.
G10L 25/78(2013.01)
(56)对比文件
CN 102667927 A,2012.09.12
CN 1168204 A,1997.12.17
审查员 李江平

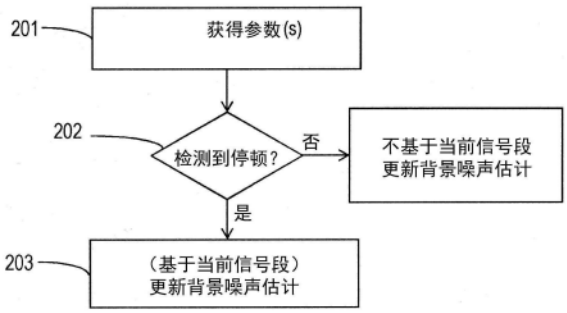
权利要求书2页 说明书30页 附图11页

(54)发明名称

用于估计背景噪声的方法和背景噪声估计器

(57)摘要

本发明涉及用于估计音频信号中的背景噪声的背景噪声估计器及其中的方法。所述方法包括基于以下各项获得与音频信号段(例如帧或帧的一部分)相关联的至少一个参数:第一线性预测增益,被计算为:针对所述音频信号段,来自0阶线性预测的残余信号与来自2阶线性预测的残余信号之间的商;以及,第二线性预测增益,被计算为:针对所述音频信号段,来自2阶线性预测的残余信号与来自16阶线性预测的残余信号之间的商。该方法还包括:至少基于所获得的至少一个参数,确定音频信号段是否包括停顿;以及,当所述音频信号段包括停顿时,基于所述音频信号段更新背景噪声估计。



CN 112927725 B

1.一种用于更新音频信号的背景噪声估计的方法,所述方法包括:

-基于以下各项获得(201)与输入音频信号段相关联的至少一个参数:

-第一线性预测增益,被计算为:针对所述音频信号段,来自第一线性预测的残余信号能量与来自第二线性预测的残余信号能量之间的商,所述第二线性预测来自比所述第一线性预测更高的阶;以及

-第二线性预测增益,被计算为:针对所述音频信号段,来自所述第二线性预测的残余信号能量与来自第三线性预测的残余信号能量之间的商,所述第三线性预测来自比所述第二线性预测更高的阶;

-至少基于所述至少一个参数,确定(202)所述音频信号段是否包括停顿;以及:

如果所述音频信号段被确定为包括停顿:

-基于所述音频信号段更新(203)背景噪声估计。

2.根据权利要求1所述的方法,其中,获得所述至少一个参数包括:

-将所述第一线性预测增益和所述第二线性预测增益限制为在预定义区间内取值。

3.根据权利要求1或2所述的方法,其中,获得所述至少一个参数包括:

-创建所述第一线性预测增益和所述第二线性预测增益中的每一个的至少一个长期估计,其中,所述长期估计还基于与至少一个在先音频信号段相关联的对应线性预测增益。

4.根据权利要求1至3中任一项所述的方法,其中,获得所述至少一个参数包括:

-确定与所述音频信号段相关联的线性预测增益之一与所述线性预测增益的长期估计之间的差。

5.根据权利要求1至4中任一项所述的方法,其中,获得所述至少一个参数包括:

确定与线性预测增益之一相关联的两个长期估计之间的差。

6.根据权利要求1至5中任一项所述的方法,其中,获得所述至少一个参数包括对所述第一线性预测增益和所述第二线性预测增益进行低通滤波。

7.根据权利要求6所述的方法,其中,至少一个低通滤波器的滤波系数取决于以下二者之间的关系:与所述音频信号段相关联的线性预测增益,以及基于多个在先音频信号段获得的对应线性预测增益的平均值。

8.根据前述权利要求中任一项所述的方法,其中,确定所述音频信号段是否包括停顿还基于与所述音频信号段相关联的谱接近度测量。

9.根据权利要求8所述的方法,还包括:基于针对所述音频信号段的频带集合和与所述频带集合相对应的背景噪声估计的能量来获得所述谱接近度测量。

10.根据权利要求9所述的方法,其中,在初始化周期期间,使用初始值E_{min}作为基于其来获得所述谱接近度测量的所述背景噪声估计。

11.一种用于更新音频信号的背景噪声估计的装置(1100),所述音频信号包括多个音频信号段,所述装置被配置为:

-基于以下各项获得至少一个参数:

-第一线性预测增益,被计算为:针对所述音频信号段,来自第一线性预测的残余信号能量与来自第二线性预测的残余信号能量之间的商,所述第二线性预测来自比所述第一线性预测更高的阶;以及

-第二线性预测增益,被计算为:针对所述音频信号段,来自所述第二线性预测的残余

信号能量与来自第三线性预测的残余信号能量之间的商,所述第三线性预测来自比所述第二线性预测更高的阶;

-至少基于所述至少一个参数,确定所述音频信号段是否包括停顿;以及

如果所述音频信号段被确定为包括停顿:

-基于所述音频信号段更新背景噪声估计。

12.根据权利要求11所述的装置,其中所述装置还被配置为执行根据权利要求2-10中任一项所述的方法。

13.一种音频编解码器,包括根据权利要求11或12所述的装置。

14.一种通信设备,包括根据权利要求11或12所述的装置。

用于估计背景噪声的方法和背景噪声估计器

[0001] 本申请是申请日为2015年7月1日,申请号为201580040591.8,题为“用于估计背景噪声的方法和背景噪声估计器”的中国专利申请的分案申请。

技术领域

[0002] 本发明的实施例涉及音频信号处理,并具体涉及背景噪声的估计,例如以支持声音活跃性判断。

背景技术

[0003] 在利用不连续传输(DTX)的通信系统中,找到效率和不降低质量之间的平衡是重要的。在这种系统中,活跃性检测器用于指示要被主动编码的活跃信号(例如语音或音乐)以及具有背景信号的段,可由在接收机侧产生的舒适噪声来替换该具有背景信号的段。如果活跃性检测器在检测不活跃中过于有效,则它将在活跃信号中引入削波,然后当被削波的活跃段被舒适噪声替换时,该活跃信号被感知为主观质量劣化。同时,如果活跃性检测器不够有效并将背景噪声段分类为活跃的,且然后主动对背景噪声进行编码,而不是进入具有舒适噪声的DTX模式,则DTX的效率降低。在多数情况下,削波问题被视为更严重。

[0004] 图1示出了广义声音活跃性检测器(SAD)或话音活跃性检测器(VAD)的概述框图,其将音频信号作为输入并产生活跃判断来作为输出。将输入信号划分为数据帧,即例如5-30ms(取决于实现)的音频信号段,并且针对每帧产生一个活跃性判断来作为输出。

[0005] 由图1中示出的主检测器做出主判断“prim”。主判断基本上只是当前帧的特征与根据先前输入帧估计的背景特征的比较。当前帧的特征与背景特征之间大于阈值的差导致活跃主判断。延迟(hangover)加法块用于基于过去的主判断扩展主判断以形成最终判断:“标志”。使用延迟的原因主要是为了减小/去除在活跃突发的中间和后端削波的风险。如图所示,操作控制器可以根据输入信号的特性调整主检测器的阈值和延迟加法的长度。使用背景估计器块来估计输入信号中的背景噪声。在本文中,背景噪声还可以被称为“背景”或者“背景特征”。

[0006] 可以根据两个基本不同的原理来进行背景特征的估计:通过使用如图1中的虚线所示的主判断(即利用判断或判断度量反馈),或者通过使用输入信号的一些其它特性(即不使用判断反馈)。还可以使用这两种策略的组合。

[0007] 使用用于背景估计的判断反馈的编解码器的示例是AMR-NB(自适应多速率窄带),且不使用判断反馈的编解码器的示例是EVRC(增强可变速率CODEC)和G.718。

[0008] 可以使用多种不同的信号特征或特性,但是在VAD中使用的一个公共特征是输入信号的频率特性。由于其低复杂度和在低SRN时的可靠操作,通常使用类型的频率特性是子带帧能量。因此假设输入信号被分割为不同频率子带,并针对每个子带估计背景电平。通过这种方式,背景噪声特征之一是具有针对每个子带的能量值的矢量,这些是在频域中表征输入信号中的背景噪声的值。

[0009] 为了实现对背景噪声的跟踪,可以通过至少三种不同方式来进行实际背景噪声估

计更新。一种方式是针对每个频率点(frequency bin)使用自动回归(AR)处理来处理更新。这种编解码器的示例是AMR-NB和G.718。基本上,对于这种类型的更新,更新的步长与观察到的当前输入和当前背景估计之间的差成正比。另一种方式是使用当前估计的乘法缩放,其限制是所述估计不能大于当前输入或者小于最小值。这意味着估计随着每帧增加,直到其高于当前输入。在此情况下,当前输入被用作估计。EVRC是使用该技术来对VAD功能的背景估计进行更新的编解码器的示例。应注意,EVRC针对VAD和噪声抑制使用不同的背景估计。应当注意的是,可以在除DTX以外的其它情境中使用VAD。例如,在可变速率编解码器(例如EVRC)中,VAD可以用作速率确定功能的一部分。

[0010] 第三种方式是使用所谓的最小技术,其中,估计是在先前帧的滑动时间窗口期间的最小值。这基本上给出了最小估计,使用补偿因子缩放该最小估计,以达到或近似针对静止噪声的平均估计。

[0011] 在高SNR情况下(其中,活跃信号的信号电平远高于背景信号),可以很容易做出输入音频信号是活跃还是非活跃的判断。然而,为了在低SNR情况下,并且尤其是当背景是非静止的或者甚至在其特征上类似于活跃信号时,对活跃和非活跃信号进行分离是非常困难的。

[0012] VAD的性能取决于背景噪声估计器跟踪背景特性的能力,尤其是在其遇到非静止背景的情况下。通过更好的跟踪,可以使得VAD更高效,而不增加语音削波的风险。

[0013] 虽然相关性是用于检测语音(主要是语音的浊音(voiced)部分)的重要特征,但也存在显示出高相关性的噪声信号。在这些情况下,具有相关性的噪声将阻止背景噪声估计的更新。结果是高活跃性,原因在于将语音和背景噪声二者均编码为活跃内容。虽然对于高SNR(大约>20dB)可以使用基于能量的停顿检测来减小该问题,但这对于从20dB降至10dB或可能降至5dB的SNR范围是不可靠的。在该范围内,本文所述的解决方案有所不同。

发明内容

[0014] 期望实现对音频信号中的背景噪声的改进估计。这里“改进的”可以意味着关于音频信号是否包括活跃语音或音乐作出更正确的判断,并因此更经常地对实际上没有活跃内容(例如语音和/或音乐)的音频信号段中的背景噪声进行估计(例如,对先前估计进行更新)。本文中,提供了一种用于生成背景噪声估计的改进方法,其可以使例如声音活跃性检测器能够做出更适当的判断。

[0015] 针对音频信号中的背景噪声估计,重要的是,当输入信号包括活跃信号和背景信号的未知混合时,还能够找到可靠特征来识别背景噪声信号的特性,其中活跃信号可以包括语音和/或音乐。

[0016] 发明人已经认识到,可以利用与针对不同线性预测模型阶数的残余能量相关的特征来检测音频信号中的停顿。可以例如从线性预测分析提取这些残余能量,这在语音编解码器中是常见的。可以对特征进行滤波和组合,以产生可以用于检测背景噪声的特征或参数集合,这使得该解决方案适合用在噪声估计中。本文描述的解决方案对于当SNR在10dB至20dB的范围中时的条件特别有效。

[0017] 本文提供的另一特征是对与背景的谱接近度的测量,其可以例如通过以下方式进行,例如通过使用例如用在子带SAD中的频域子带能量。谱接近度测量还可以用于做出音频

信号是否包括停顿的判断。

[0018] 根据第一方面,提供了一种用于背景噪声估计的方法。所述方法包括基于以下各项获得与音频信号段(例如帧或帧的一部分)相关联的至少一个参数:第一线性预测增益,被计算为:针对所述音频信号段,来自0阶线性预测的残余信号与来自2阶线性预测的残余信号之间的商;以及,第二线性预测增益,被计算为:针对所述音频信号段,来自2阶线性预测的残余信号与来自16阶线性预测的残余信号之间的商。该方法还包括:至少基于所获得的至少一个参数,确定音频信号段是否包括停顿;以及,当所述音频信号段包括停顿时,基于所述音频信号段更新背景噪声估计。

[0019] 根据第二方案,提供了一种背景噪声估计器。背景噪声估计器被配置为:基于以下各项获得与音频信号段相关联的至少一个参数:第一线性预测增益,被计算为:针对所述音频信号段,来自0阶线性预测的残余信号与来自2阶线性预测的残余信号之间的商;以及,第二线性预测增益,被计算为:针对所述音频信号段,来自2阶线性预测的残余信号与来自16阶线性预测的残余信号之间的商。背景噪声估计器还被配置为:至少基于所述至少一个参数,确定所述音频信号段是否包括停顿;以及,当所述音频信号段包括停顿时,基于所述音频信号段更新背景噪声估计。

[0020] 根据第三方面,提供了一种SAD,其包括根据第二方面的背景噪声估计器。

[0021] 根据第四方面,提供了一种编解码器,其包括根据第二方面的背景噪声估计器。

[0022] 根据第五方面,提供了一种通信设备,其包括根据第二方面的背景噪声估计器。

[0023] 根据第六方面,提供了一种网络节点,其包括根据第二方面的背景噪声估计器。

[0024] 根据第七方面,提供了一种计算机程序,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方面的方法。

[0025] 根据第八方面,提供了一种载体,其包含根据第七方面的计算机程序。

附图说明

[0026] 根据以下对附图中示出的实施例更具体的描述,本文公开的技术的以上或其他目的、特征、优点将显而易见。附图不必须按比例绘制,相反重点在于说明本文公开的技术的原理。

[0027] 图1是示出活跃性检测器和延迟确定逻辑的框图。

[0028] 图2是示出根据示例实施例的用于估计背景噪声的方法的流程图。

[0029] 图3是示出根据示例性实施例的特征计算的框图,该特征与针对阶数为0和2的线性预测的残余能量相关。

[0030] 图4是示出根据示例性实施例的特征计算的框图,该特征与针对阶数为2和16的线性预测的残余能量相关。

[0031] 图5是示出根据示例性实施例的特征计算的框图,该特征与谱接近度测量相关。

[0032] 图6是示出子带能量背景估计器的框图。

[0033] 图7是示出来自附录A中描述的解决方案的背景更新判断逻辑的流程图。

[0034] 图8-图10是示出当针对包括两个语音突发的音频信号进行计算时本文所呈现的不同参数的表现的示图。

[0035] 图11a-图11c和图12-图13是示出根据示例实施例的背景噪声估计器的不同实现

的框图。

[0036] 图14是示出了根据本文提出的技术的用于背景噪声估计的方法的示例性实施例的流程图。

[0037] 图15是示意性示出了根据示例性实施例的背景估计器的实现的框图。

具体实施方式

[0038] 本文公开的方案涉及估计音频信号中的背景噪声。在图1中所示的广义活跃性检测器中,由被表示为“背景估计器”的块来执行估计背景噪声的功能。可以在之前公开的WO2011/049514、WO2011/049515(其通过引用并入本文)的解决方案中以及在附录A(附录A)中找到与本方案相关的一些实施例。本文公开的解决方案将与这些先前公开的解决方案的实现进行比较。即使WO2011/049514、WO2011/049515和附录A中公开的解决方案是优良的解决方案,但是本文呈现的解决方案仍然具有相对于这些解决方案的优点。例如,本文呈现的解决方案更胜任其对背景噪声的跟踪。

[0039] VAD的性能取决于背景噪声估计器跟踪背景特性的能力,尤其是在其遇到非静止背景的情况下。通过更好的跟踪,可以使得VAD更高效,而不增加语音削波的风险。

[0040] 当前噪声估计方法的一个问题在于为了实现在低SNR中对背景噪声进行良好跟踪,需要可靠的停顿检测器。对于只有语音的输入,可以利用音节速率或者人不可能一直说话的事实来找到语音中的停顿。这种方案可以涉及在不进行背景更新的足够时间之后,“放松”对停顿检测的需要,从而更可能检测语音中的停顿。这允许对噪声特性或电平中的突然改变进行响应。这种噪声恢复逻辑的一些示例是:1)由于语音话语包含具有高相关性的段,假定在足够数量的无相关性的帧之后在语音中存在停顿,这通常是安全的。2)当信噪比SNR>0时,语音能量高于背景噪声,所以如果帧能量在长时间(例如1-5秒)内接近最小能量,假设这在语音停顿中也是安全的。虽然先前的技术对于只有语音的输入工作良好,但是当将音乐视为是活跃输入时它们是不够的。在音乐中,可能存在具有低相关性的长段,但是其仍是音乐。此外,音乐中的能量的动态性也可以触发假停顿检测,这可能导致对背景噪声估计的不想要的错误更新。

[0041] 理想地,将需要活跃性检测器的反功能(或将被称为“停顿发生检测器”),以控制噪声估计。这将确保仅在当前帧中没有活跃信号时才进行背景噪声特性的更新。然而,如上所述,确定音频信号段是否包括活跃信号并不容易。

[0042] 传统地,当已知活跃信号是语音信号时,活跃性检测器被称为语音活跃性检测器(VAD)。当输入信号可以包括音乐时,也经常使用用于活跃性检测器的术语VAD。然而,在现代编解码器中,当还要将音乐作为活跃信号检测时,将活跃性检测器称为声音活跃性检测器(SAD)也是常见的。

[0043] 图1中所示的背景估计器利用来自主检测器的反馈和/或延迟块来定位不活跃的音频信号段。当开发本文所述的技术时,期望移除或者至少减少对这种反馈的依赖性。对于本文公开的背景估计,因此发明人已经认识到,当只有具有活跃信号和背景信号的未知混合的输入信号可用时,能够找到可靠特征来识别背景信号特性是重要的。发明人还认识到,不能假定输入信号以噪声段开始,或者甚至输入信号是与噪声混合的语音,原因在于活跃信号可能是音乐。

[0044] 一个方案是,即使当前帧可能具有与当前噪声估计相同的能量电平(level),频率特性也可能差异很大,这使得不期望使用当前帧来执行噪声估计的更新。所引入的接近度特征相关背景噪声更新可以用于防止在这些情况下更新。

[0045] 此外,在初始化期间,期望允许噪声估计尽可能快地开始,同时避免错误的判断,原因在于如果使用活跃内容来进行背景噪声更新,则这潜在地可能导致来自SAD的削波。在初始化期间使用接近度特征的初始化特定版本可以至少部分解决该问题。

[0046] 本文描述的解决方案涉及一种用于背景噪声估计的方法,具体涉及一种在困难的SNR情况下良好执行的用于检测音频信号中的停顿的方法。下面将参照图2-5描述该解决方案。

[0047] 在语音编码领域中,通常使用所谓的线性预测来分析输入信号的谱形状。通常每帧进行两次分析,并为了提高的时间精度,然后对结果进行内插,使得可针对输入信号的每个5ms的块生成滤波。

[0048] 线性预测是一种数学运算,其中离散时间信号的未来值被估计为先前采样的线性函数。在数字信号处理中,线性预测通常被称为线性预测编码(LPC),并因此可以被视为滤波器理论的子集。在语音编码器的线性预测中,将线性预测滤波器A(z)应用于输入语音信号。A(z)是全零滤波器,其在应用于输入信号时,从输入信号中去除可以使用滤波器A(z)来建模的冗余。因此,当滤波器对输入信号的某一方面或某些方面成功建模时,滤波器的输出信号具有比输入信号更低的能量。该输出信号表示为“残差”、“残余能量”或“残余信号”。这种线性预测滤波器(备选地表示为残余滤波器)可以具有的不同模型阶数,该的不同模型阶数具有不同数量的滤波系数。例如,为了对语音合适地建模,可能需要模型阶数为16的线性预测滤波器。因此,在语音编码器中,可以使用模型阶数为16的线性预测滤波器A(z)。

[0049] 发明人已经认识到,与线性预测相关的特征可以用于检测在20dB降至10dB或可能降至5dB的SNR范围内的音频信号中的停顿。根据本文描述的解决方案的实施例,使用针对音频信号的不同模型阶数的残余能量之间的关系来检测音频信号中的停顿。所使用的关系是较低模型阶数和较高模型阶数的残余能量之间的商。残余能量之间的商可以被称为“线性预测增益”,原因在于它是对线性预测滤波器在一个模型阶数与另一模型阶数之间能够对多少信号能量进行建模或能够去除多少信号能量的指示符。

[0050] 残余能量将取决于线性预测滤波器A(z)的模型阶数M。计算线性预测滤波器的滤波系数的常用方法是Levinson-Durbin算法。该算法是递归的,并且还将在创建阶数为M的预测滤波器A(z)的过程中,作为“副产品”产生较低模型阶数的残余能量。根据本发明的实施例,可以利用该事实。

[0051] 图2示出了用于估计音频信号中的背景噪声的示例性一般方法。可以由背景噪声估计器执行该方法。该方法包括:基于以下各项获得201与音频信号段(例如帧或帧的一部分)相关联的至少一个参数:第一线性预测增益,被计算为:针对音频信号段,来自0阶线性预测的残余信号与来自2阶线性预测的残余信号之间的商;以及,第二线性预测增益,被计算为:针对音频信号段,来自2阶线性预测的残余信号与来自16阶线性预测的残余信号之间的商。

[0052] 该方法还包括:至少基于所获得的至少一个参数,确定202音频信号段是否包括暂停,即没有诸如语音和音乐的活跃内容;以及,当音频信号段包括停顿时,基于该音频信号

段更新背景噪声估计。也就是说,该方法包括:当至少基于所获得的至少一个参数在音频信号段中检测到停顿时,更新背景噪声估计。

[0053] 线性预测增益可以被描述为与从音频信号段的0阶到2阶线性预测相关的第一线性预测增益;以及与从音频信号段的2阶到16阶线性预测相关的第二线性预测增益。此外,至少一个参数的获得可备选地被描述为确定、计算、导出或创建。可以从编码器的在其处将线性预测作为常规编码处理的一部分来执行的部分获得、接收或检索(即由某种方式提供)与模型阶数0、2和16的线性预测相关的残余能量。由此,与当需要导出残余能量(尤其是针对估计背景噪声)时相比,可以减小本文所述的解决方案的计算复杂度。

[0054] 基于线性预测特征获得的至少一个参数可以提供输入信号的电平独立分析,其改进了对是否执行背景噪声更新的判断。该解决方案在10至20dB的SNR范围内特别有用,在该范围内,基于能量的SAD由于语音信号的正常动态范围而具有有限的性能。

[0055] 在本文中,变量E(0),...,E(m),...,E(M)表示针对M+1个滤波器Am(z)的模型阶数0到M的残余能量。注意E(0)正是输入能量。根据本文所述的解决方案的音频信号分析通过分析以下线性预测增益来提供若干新特征或参数:被计算为来自0阶线性预测的残余信号与来自2阶线性预测的残余信号之间的商的线性预测增益,以及被计算为来自2阶线性预测的残余信号与来自16阶线性预测的残余信号之间的商的线性预测增益。也就是说,针对从0阶线性预测到2阶线性预测的线性预测增益与残余能量E(0)(针对第0模型阶数)除以残余能量E(2)(针对第2模型阶数)是同一回事。对应地,针对从2阶线性预测到16阶线性预测的线性预测增益与残余能量E(2)(针对第二模型阶数)除以残余能量E(16)(针对第16模型阶数)是同一回事。将在下面进一步更详细地描述参数的示例和基于预测增益确定参数。根据上述一般实施例获得的至少一个参数可以形成用于评估是否更新背景噪声估计的判断准则的一部分。

[0056] 为提高该至少一个参数或特征的长期稳定性,可以计算预测增益的受限版本。也就是说,获得至少一个参数可以包括:将与从0阶到2阶以及从2阶到16阶线性预测相关的线性预测增益限制为在预定义区间内取值。例如,如例如在以下等式1和等式6中所指示的,线性预测增益可以被限制为取0和8之间的值。

[0057] 获得至少一个参数还可包括:例如通过低通滤波的方式来创建所述第一线性预测增益和所述第二线性预测增益中的每一个的至少一个长期估计。这种至少一个长期估计然后还将基于与至少一个在先音频信号段相关联的对应线性预测增益。可以创建多于一个长期估计,其中例如与线性预测增益相关的第一和第二长期估计对音频信号的变化作出不同反应。例如,与第二长期估计相比,第一长期估计可以对变化反应更快。这种第一长期估计可备选地被表示为短期估计。

[0058] 获得至少一个参数还可以包括:确定与音频信号段相关联的线性预测增益之一与所述线性预测增益的长期估计之间的差,例如下面描述的绝对差Gd_0_2(等式3)。备选地或此外,可以确定两个长期估计之间的差,例如在以下等式9中。术语“确定”可以备选地与计算、创建或导出交换。

[0059] 获得至少一个参数可以如上所述包括:线性预测增益的低通滤波,从而导出长期估计,其中一些可以备选地被表示为短期估计,这取决于在估计中考虑多少个段。至少一个低通滤波器的滤波系数可以取决于(例如仅)与当前信号段相关的线性预测增益与基于多

个在先音频信号段获得的对应预测增益的平均值(表示为例如长期平均值)或长期估计之间的关系。可以对此进行执行以创建预测增益的例如进一步长期估计。低通滤波可以在两个或更多个步骤中执行,其中每个步骤可以产生用于做出与音频信号段中停顿的存在有关的判断的参数或估计。例如,可以对用不同的方式反映音频信号中的变化的不同长期估计(例如下面描述的G1_0_2(等式2)和Gad_0_2(等式4)和/或G1_2_16(等式7)、G2_2_16(等式8)和Gad_2_16)进行分析或比较,以检测当前音频信号段中的停顿。

[0060] 确定202音频信号段是否包括停顿还可以基于与音频信号段相关联的谱接近度测量。谱接近度测量将指示当前处理的音频信号段的“每频带”能量电平与当前背景噪声估计的“每频带”能量电平(例如,作为在对当前音频信号段进行分析之前进行的先前更新的结果的初始值或估计)的接近程度。以下在等式等式12和等式13中给出确定或导出谱接近度测量的示例。谱接近度测量可以用于防止基于低能量帧的噪声更新,该低能量帧与当前背景估计相比具有较大差异的频率特征。例如,针对当前信号段和当前背景噪声估计,频带上的平均能量可以同样低,但谱接近度测量将揭示能量是否在频带上不同地分布。这种能量分布的差异可以表明当前信号段(例如,帧)可以是低电平活跃内容,且基于该帧的背景噪声估计更新可例如防止检测具有相似内容的未来帧。由于子带SNR对能量增加最敏感,如果在背景噪声中不存在该特定频率范围(例如,与低频车辆噪声相比的语音的高频部分),则使用甚至低电平活跃内容也可导致背景估计的较大更新。在这种更新之后,将更难以检测语音。

[0061] 如上所述,可以基于针对当前分析的音频信号段的频带集合(备选地表示为子带)和与该频率集合相对应的当前背景噪声的能量估计来导出、获得或计算谱接近度测量。这也将在下面进一步更详细地示例和描述,并在图5中示出。

[0062] 如上所述,可以通过将当前处理的音频信号段的当前每频带能量电平与当前背景噪声估计的每频带能量电平进行比较,来导出、获得或计算谱接近度测量。然而,开始时(即在开始分析音频信号时的第一周期或第一数量的帧期间),可能没有可靠的背景噪声估计,例如原因在于尚未执行背景噪声估计的可靠更新。因此,可以应用初始化周期来确定谱接近度值。在这样的初始化周期期间,当前音频信号段的每频带能量电平将代之以与初始背景估计进行比较,初始背景估计可以是例如可配置的恒定值。在更下面的示例中,该初始背景噪声估计被设置为示例值$E_{min}=0.0035$。在初始化周期之后,该过程可以切换为正常操作,并将当前处理的音频信号段的当前每频带能量电平与当前背景噪声估计的每频带能量电平进行比较。初始化周期的长度可以例如基于模拟或测试来配置,其指示在例如提供可靠和/或令人满意的背景噪声估计之前所花费的时间。下面使用的示例,在前150个帧期间执行与初始背景噪声估计(而不是与基于当前音频信号导出的“真实”估计)的比较。

[0063] 该至少一个参数可以是在下面的代码中例示的参数(表示为NEW_POS_BG)和/或下面进一步描述的多个参数中的一个或多个,导致形成用于停顿检测的判断准则或判断准则中的组成部分。换言之,基于线性预测增益获得201的至少一个参数或特征可以是下面描述的参数中的一个或多个、可以包括下面描述的参数中的一个或多个和/或基于以下描述的参数中的一个或多个。

[0064] 与残余能量E(0)和E(2)相关的特征或参数

[0065] 图3示出根据示例性实施例的导出与E(0)和E(2)相关的特征或参数的概述框图。

从图3中可以看出,预测增益首先被计算为E(0)/E(2)。预测增益的受限版本计算为

[0066] G_0_2＝max(0,min(8,E(0)/E(2))) (等式1)

[0067] 其中E(0)表示输入信号的能量,并且E(2)是2阶线性预测之后的残余能量。等式1中的表达式将预测增益限制在0和8之间的区间。在正常情况下预测增益应当大于零,但例如针对接近零的值可能发生异常,并且因此“大于零”限制(0<)可能是有用的。将预测增益限制到最大值为8的原因在于:为了本文所描述的解决方案的目的,知道预测增益为大约8或大于8(其指示显著的线性预测增益)就足矣。应当注意,当两个不同模型阶数之间的残余能量之间没有差异时,线性预测增益将为1,这指示较高模型阶数的滤波器在对音频信号建模中与较低模型阶数的滤波器相比并不更为成功。此外,如果预测增益G_0_2在下面的表达式中取过大的值,则可能关于所导出的参数的稳定性存在风险。应当注意,8仅是针对特定实施例选择的示例值。参数G_0_2可以备选地表示为例如epsP_0_2或$g_{LP_{0_2}}$。

[0068] 然后在两个步骤中对受限的预测增益进行滤波以创建该增益的长期估计。第一低通滤波以及因此第一长期特征或参数的导出如下进行:

[0069] G1_0_2＝0.85G1_0_2+0.15G_0_2, (等式2)

[0070] 其中表达式中的第二个“G1_0_2”应被读取为来自在先音频信号段的值。一旦存在仅具有背景的输入的段,取决于输入中的背景噪声的类型,该参数通常会是0或8。参数G1_0_2可以备选地表示为例如epsP_0_2_lp或$\bar{g}_{LP_0_2}$。然后可以根据以下等式使用第一长期特征G1_0_2与逐帧限制的预测增益G_0_2之间的差来创建或计算另一特征或参数:

[0071] Gd_0_2＝abs(G1_0_2-G_0_2) (等式3)

[0072] 这将给出与预测增益的长期估计相比的当前帧的预测增益的指示。参数Gd_0_2可以备选地表示为例如epsP_0_2_ad或$g_{ad_0_2}$。在图4中,该差被用于创建第二长期估计或特征Gad_0_2。这是根据依据以下等式使用应用不同滤波系数的滤波器来完成的,该不同滤波系数取决于长期差异是高于还是低于当前估计的平均差异:

[0073] Gad_0_2＝(1-a)Gad_0_2+a Gd_0_2 (等式4)

[0074] 其中,如果Gd_0_2<Gad_0_2,则a＝0.1否则a＝0.2

[0075] 其中表达式中的第二个“Gad_0_2”应被读取为来自在先音频信号段的值。

[0076] 参数Gad_0_2可以备选地表示为例如Glp_0_2、epsP_0_2_ad_lp或$\bar{g}_{ad_0_2}$。为防止过滤屏蔽偶尔的高的帧差异,可以导出另一参数,其在图中未示出。也就是说,第二长期特征Gad_0_2可以与帧差异组合,以防止这种屏蔽。可以如下通过取预测增益特征的帧版本Gd_0_2和长期版本Gad_0_2中的最大值来导出该参数:

[0077] Gmax_0_2＝max(Gad_0_2,Gd_0_2) (等式5)

[0078] 参数Gmax_0_2可以备选地表示为epsP_0_2_ad_lp_max或$g_{max_0_2}$。

[0079] 与残余能量E(2)和E(16)相关的特征或参数

[0080] 图4示出根据示例性实施例的导出与E(2)和E(16)相关的特征或参数的概述框图。从图4中可以看出,预测增益首先被计算为E(2)/E(16)。与以上关于0阶和2阶残余能量之间的关系描述的特征或参数稍有不同地导出使用2阶残余能量和16阶残余能量之间的差异或关系创建的特征或参数。

[0081] 这里,受限的预测增益还被计算为:

[0082] G_2_16＝max(0,min(8,E(2)/E(16))) (等式6)

[0083] 其中E(2)表示2阶线性预测之后的残余能量,并且E(16)表示16阶线性预测之后的残余能量。参数G_2_16可以备选地表示为例如epsP_2_16或$g_{LP_2_16}$。然后该受限的预测增益被用于创建该增益的两个长期估计:在一个长期估计中,不管长期估计要增加与否,滤波系数不同,如下所示:

[0084] G1_2_16＝(1-a)G1_2_16+a G_2_16 (等式7)

[0085] 其中,如果G_2_16>G1_2_16,则a＝0.2,否则a＝0.03

[0086] 参数G1_2_16可以备选地表示为例如epsP_2_16_lp或$\bar{g}_{LP_2_16}$。

[0087] 第二长期估计使用根据以下等式的恒定滤波系数:

[0088] G2_2_16＝(1-b)G2_2_16+b G_2_16,其中b＝0.02 (等式8)

[0089] 参数G2_2_16可以备选地表示为例如epsP_2_16_lp2或$\bar{g}_{LP2_0_2}$。

[0090] 对于大多数类型的背景信号,G1_2_16和G2_2_16均将接近0,但是它们对于需要16阶线性预测的内容具有不同的响应,这通常针对于语音和其他活跃内容而言。第一长期估计G1_2_16通常将高于第二长期估计G2_2_16。根据以下等式测量长期特征之间的该差异:

[0091] Gd_2_16＝G1_2_16-G2_2_16 (等式9)

[0092] 参数Gd_2_16可以备选地表示为epsP_2_16_dlp,或者$g_{ad_2_16}$。

[0093] Gd_2_16然后可以用作滤波器的输入,该滤波器根据以下等式创建第三长期特征:

[0094] Gad_2_16＝(1-c)Gad_2_16+c Gd_2_16 (等式10)

[0095] 其中,如果Gd_2_16<Gad_2_16,则c＝0.02,否则c＝0.05

[0096] 该滤波器根据第三长期信号是否增长来应用不同的滤波系数。参数Gad_2_16可以备选地表示为例如epsP_2_16_dlp_lp2或$\bar{g}_{ad_2_16}$。这里,长期信号Gad_2_16可以与滤波器输入信号Gd_2_16组合,以防止滤波屏蔽针对当前帧的偶尔的高输入。从而,最终参数是帧或段以及特征的长期版本中的最大值

[0097] Gmax_2_16＝max(Gad_2_16,Gd_2_16) (等式11)

[0098] 参数Gmax_2_16可以备选地表示为例如epsP_2_16_dlp_max或$g_{max_0_2}$。

[0099] 谱接近度/差异测量

[0100] 谱接近度特征使用当前输入帧或段的频率分析,其中计算子带能量并将其与子带背景估计进行比较。谱接近度参数或特征可以与上述与线性预测增益相关的参数组合使用,例如以确保当前段或帧相对接近于或至少不过于远离先前背景估计。

[0101] 图5示出了谱接近度或差异测量的计算的框图。在初始化周期期间,例如在前150帧中,与对应于初始背景估计的常数进行比较。在初始化之后,进入正常操作并与背景估计进行比较。注意,虽然谱分析产生20个子带的子带能量,但是这里的nonstaB的计算仅使用子带i＝2,...16,原因在于它主要在语音能量所位于的这些频带中。这里nonstaB反映了非静止性。

[0102] 因此,在初始化期间,使用Emin(这里将Emin设置为Emin＝0.0035)来计算nonstaB:

[0103] nonstaB＝sum(abs(log(Ecb(i)+1)-log(Emin+1))) (等式12)

[0104] 其中,在i＝2...16上求和。

[0105] 这样做是为了减小初始化期间背景噪声估计中判断误差的影响。在初始化周期之后,根据下式使用对应子带的当前背景噪声估计进行计算:

[0106] nonstaB＝sum(abs(log(Ecb(i)+1)-log(Ncb(i)+1))) (等式13)

[0107] 其中,在i＝2...16上求和。

[0108] 在对数之前将常数1添加到每个子带能量降低了对低能量帧的谱差的灵敏度。参数nonstaB可以备选地表示为例如non_staB或$nonstat_B$。

[0109] 图6中示出了说明背景估计器的示例性实施例的框图。图6中的实施例包括用于输入成帧601的块,其将输入音频信号划分为适当长度(例如,5-30毫秒)的帧或段。该实施例还包括用于特征提取602的块,其计算针对输入信号的每个帧或段的特征(在本文中也表示为参数)。该实施例还包括用于更新判断逻辑603的块,用于确定是否可以基于当前帧中的信号(即,信号段是否没有诸如语音和音乐的活跃内容)来更新背景估计。该实施例还包括背景更新器604,用于当更新判断逻辑指示更新背景噪声估计是适当的时对背景噪声估计进行更新。在所示实施例中,可以针对每个子带(即针对多个频带)导出背景噪声估计。

[0110] 本文所描述的解决方案可用于改进如本文附录A以及文献WO2011/049514中所描述的用于背景噪声估计的先前的解决方案。下面,将在上述解决方案的上下文中描述本文描述的解决方案。将给出来自背景噪声估计器的实施例的代码实现的代码示例。

[0111] 下面,在基于G.718的编码器中针对本发明的实施例描述实际实现细节。该实现使用在附录A和通过引用并入本文的WO2011/049514中的解决方案中描述的许多能量特征。参考附录A和WO2011/049514以寻求比以下所呈现的细节更多的进一步细节。

[0112] WO2011/049514中定义了以下能量特征:

[0113] Etot;

[0114] Etot_l_lp;

[0115] Etot_v_h;

[0116] totalNoise;

[0117] sign_dyn_lp;

[0118] WO2011/049514中定义了以下相关性特征:

[0119] aEn;

[0120] harm_cor_cnt

[0121] act_pred

[0122] cor_est

[0123] 在附录A中给出的解决方案中定义了以下特征:

```
Etot_v_h;
lt_cor_est = 0.01f*cor_est + 0.99f*lt_cor_est;
lt_tn_track = 0.03f* (Etot - totalNoise < 10) + 0.97f*lt_tn_track;
```

[0124]

```
lt_tn_dist = 0.03f* (Etot - totalNoise) + 0.97f*lt_tn_dist;
lt_Ellp_dist = 0.03f* (Etot - Etot_l_lp) + 0.97f*lt_Ellp_dist;
harm_cor_cnt
low_tn_track_cnt
```

[0125] 在附图7中示出了来自附录A中给出的解决方案的噪声更新逻辑。对附录A的噪声估计器的与本文所述的解决方案相关的改进主要涉及以下部分:计算特征的部分701;部分702,其中基于不同的参数作出停顿判断;以及还涉及部分703,其中基于是否检测到停顿来采取不同的动作。此外,这些改进可能对背景噪声估计的更新704造成影响,可例如当基于新特征检测到在引入本文所描述的解决方案之前可能检测不到的停顿时更新背景噪声估计。在这里所描述的示例性实现中,如下计算本文所引入的新特征,以non_staB开始,其是使用与以上和图6中的Ecb(i)相对应的当前帧的子带能量enr[i]以及与以上和图6中的Ncb(i)相对应的当前背景噪声估计bckr[i]来确定的。下面的第一代码段的第一部分与在导出适当的背景估计之前音频信号的前150个帧的特定初始过程相关。

```
/*计算非静止性特征相关背景（谱接近度特征 non_staB */
if (ini_frame < 150)
{
    /*在初始化期间不包括更新*/
    if ( i >= 2 && i <= 16 )
    {
        non_staB += (float)fabs(log(enr[i] + 1.0f) -
                                log(E_MIN + 1.0f));
    }
}
else
{
    /*在初始化之后与背景估计进行比较*/
    if ( i >= 2 && i <= 16 )
    {
        non_staB += (float)fabs(log(enr[i] + 1.0f) -
                                log(bckr[i] + 1.0f));
    }
}
```

[0126]

```
if (non_staB >= 128)
{
    non_staB = 32767.0/256.0f;
}
```

[0127]

[0128] 下面的代码段示出了如何计算线性预测残余能量(即针对于线性预测增益)的新特征。这里,残余能量被命名为epsP[m](参考先前使用的E(m))。

[0129]

```
/*---------------------------------------------------------------*
*线性预测效率 0 至 2 阶
*（从线性预测滤波器的 0 阶模型至 2 阶模型的线性预测增益）
/*---------------------------------------------------------------*
epsP_0_2 = max(0 , min(8, epsP[0] / epsP[2]));
epsP_0_2_lp = 0.15f * epsP_0_2 + (1.0f-0.15f) * st->epsP_0_2_lp;
epsP_0_2_ad = (float) fabs(epsP_0_2 - epsP_0_2_lp );
if (epsP_0_2_ad < epsP_0_2_ad_lp)
{
    epsP_0_2_ad_lp = 0.1f * epsP_0_2_ad + (1.0f - 0.1f) * epsP_0_2_ad_lp;
}
else
{
    epsP_0_2_ad_lp = 0.2f * epsP_0_2_ad + (1.0f - 0.2f) * epsP_0_2_ad_lp;
}
epsP_0_2_ad_lp_max = max(epsP_0_2_ad,st->epsP_0_2_ad_lp);

/*---------------------------------------------------------------*
*线性预测效率 2 至 16 阶
*（线性预测滤波器的 2 至 16 阶模型的线性预测增益）
```

```
/*-------------------------------------------------------------*

epsP_2_16 = max(0 , min(8, epsP[2] / epsP[16]));
if (epsP_2_16 > epsP_2_16_lp)
{
    epsP_2_16_lp = 0.2f * epsP_2_16 + (1.0f-0.2f) * epsP_2_16_lp;
}
else
{
    epsP_2_16_lp = 0.03f * epsP_2_16 + (1.0f-0.03f) * epsP_2_16_lp;
}
epsP_2_16_lp2 = 0.02f * epsP_2_16 + (1.0f-0.02f) * epsP_2_16_lp2;
```

[0130]

```
epsP_2_16_dlp = epsP_2_16_lp-epsP_2_16_lp2;

if (epsP_2_16_dlp < epsP_2_16_dlp_lp2 )
{
    epsP_2_16_dlp_lp2 = 0.02f * epsP_2_16_dlp + (1.0f-0.02f) *
epsP_2_16_dlp_lp2;
}
else
{
    epsP_2_16_dlp_lp2 = 0.05f * epsP_2_16_dlp + (1.0f-0.05f) *
epsP_2_16_dlp_lp2;
}

    epsP_2_16_dlp_max = max(epsP_2_16_dlp,epsP_2_16_dlp_lp2);
```

[0131] 以下代码示出了对用于实际更新判断(即,确定是否更新背景噪声估计)的组合度量、阈值和标志的创建。以粗体文本指示与线性预测增益和/或谱接近度相关的参数中的至少一些。

comb_ahc_epsP = max(max(act_pred,lt_haco_ev),**epsP_2_16_dlp**);

comb_hcm_epsP= max(max(lt_haco_ev,**epsP_2_16_dlp_max**),**epsP_0_2_ad_lp_max**);

haco_ev_max = max(st_harm_cor_cnt==0,>lt_haco_ev);

Etot_l_lp_thr = st->Etot_l_lp + (1.5f + 1.5f * (Etot_lp<50.0f))*Etot_v_h2;

enr_bgd = Etot < Etot_l_lp_thr;

cns_bgd = (**epsP_0_2** > 7.95f) && (non_sta< 1e3f);

lp_bgd = **epsP_2_16_dlp_max** < 0.10f;

ns_mask = non_sta < 1e5f;

lt_haco_mask = lt_haco_ev < 0.5f;

bg_haco_mask = haco_ev_max < 0.4f;

[0132] SD_1 = ((**epsP_0_2_ad** > 0.5f) && (**epsP_0_2** > 7.95f));

bg_bgd3 = enr_bgd || ((cns_bgd || lp_bgd) && ns_mask && lt_haco_mask && SD_1==0);

PD_1 = (**epsP_2_16_dlp_max** < 0.10f) ;

PD_2 = (**epsP_0_2_ad_lp_max** < 0.10f) ;

PD_3 = (comb_ahc_epsP < 0.85f);

PD_4 = comb_ahc_epsP < 0.15f;

PD_5 = comb_hcm_epsP < 0.30f;

BG_1 = ((SD_1==0) || (Etot < Etot_l_lp_thr)) && bg_haco_mask && (act_pred < 0.85f) && (Etot_lp < 50.0f);

PAU = (aEn==0) || ((Etot < 55.0f) && (SD_1==0) && ((PD_3 && (PD_1 || PD_2)) ||

```
( PD_4 || PD_5 ) ) );

NEW_POS_BG = (PAU | BG_1) & bg_bgd3;

/*原始静音检测器在大多数情况下起作用*/
aE_bgd = aEn == 0;

/*当信号动态性高并能量接近背景估计时*/
sd1_bgd = (st->sign_dyn_lp > 15) && (Etot - st->Etot_l_lp ) < 2*st->Etot_v_h2 &&
st->harm_cor_cnt > 20;
```

[0133]

```
/* 初始条件稳定地丢弃（drop）act_pred 和/或 lt_haco_ev */
tn_ini = ini_frame < 150 && harm_cor_cnt > 5 &&
                    ( (st->act_pred < 0.59f && st->lt_haco_ev <0.23f ) ||
                      st->act_pred < 0.38f ||
                      st->lt_haco_ev < 0.15f ||
                      non_staB < 50.0f ||
                      aE_bgd );
  /*接近背景估计的能量用作其他背景检测器的掩码*/
  bg_bgd2 = Etot < Etot_l_lp_thr || tn_ini ;
```

[0134] 由于重要的是在当前帧或段包括活跃内容时不进行背景噪声估计的更新,评估若干条件以判断是否做出更新。噪声更新逻辑中的主判断步骤是是否进行更新,并这是通过对以下加下划线的逻辑表达式的评估而形成的。新参数NEW_POS_BG(相对于附录A和WO 2011/049514中的解决方案是新的)是停顿检测器,且是基于线性预测滤波器的从0阶至2阶模型以及从2阶至第16阶模型的线性预测增益而获得的,并且tn_ini是基于与谱接近度相关的特征而获得的。下面是根据示例性实施例的使用新特征的判断逻辑。

[0135]

```
updt_step=0.0f;
    if (( bg_bgd2 && ( aE_bgd || sd1_bgd || lt_tn_track >0.90f || NEW_POS_BG ) ) ||
          tn_ini )
    {
      if( ( ( act_pred < 0.85f )   &&
              aE_bgd &&
            ( lt_Ellp_dist < 10   || sd1_bgd ) && lt_tn_dist<40 &&
            ( ( Etot - totalNoise ) < 10.0f ) ) ||
          ( st->first_noise_updt == 0 && st->harm_cor_cnt > 80 && aE_bgd &&
st->lt_aEn_zero > 0.5f ) ||
          ( tn_ini && ( aE_bgd || non_staB < 10.0 || st->harm_cor_cnt > 80 ) )
        )
        {
            updt_step=1.0f;
            st->first_noise_updt = 1;
            for( i=0; i< NB_BANDS; i++ )
            {
                st->bckr[i] = tmpN[i];
            }
        }
      else if ( ( ( st->act_pred < 0.80f ) && ( aE_bgd || PAU )   &&   st->lt_haco_ev <
0.10f )   ||
                ( ( st->act_pred < 0.70f ) && ( aE_bgd || non_staB < 17.0f ) && PAU
&&   st->lt_haco_ev < 0.15f ) ||
                ( st->harm_cor_cnt > 80 && st->totalNoise > 5.0f && Etot <
max(1.0f,Etot_l_lp + 1.5f* st->Etot_v_h2) ) ||
                ( st->harm_cor_cnt > 50 && st->first_noise_updt > 30 && aE_bgd
&& st->lt_aEn_zero>0.5f ) ||
                tn_ini
              )
```

[0136]

```
        {
          updt_step=0.1f;
          if ( !aE_bgd &&
                st->harm_cor_cnt < 50 &&
                      ( st->act_pred > 0.6f ||
                        ( !tn_ini && Etot_l_lp - st->totalNoise < 10.0f &&
non_staB > 8.0f ) ) )
          {
              updt_step=0.01f;
          }
          if (updt_step > 0.0f )
            {
              st->first_noise_updt = 1;
              for( i=0; i< NB_BANDS; i++ )
              {
                  st->bckr[i] = st->bckr[i] + updt_step * (tmpN[i]-st->bckr[i]);
              }
            }
        }
        else if (aE_bgd || st->harm_cor_cnt > 100 )
        {
          (   st->first_noise_updt) += 1;
        }
    }
    else
    {
        /*如果在音乐中，进一步降低较低的bckr */
        if ( st->low_tn_track_cnt > 300 && st->lt_haco_ev >0.9f && st->totalNoise >
0.0f)
        {
```

```
                updt_step=-0.02f;
                for( i=0; i< NB_BANDS; i++ )
                {
                    if (st->bckr[i] > 2*E_MIN)
                    {
[0137]                  st->bckr[i] = 0.98f*st->bckr[i];
                    }
                }
            }
        }
        st->lt_aEn_zero = 0.2f * (st->aEn==0) + (1-0.2f)*st->lt_aEn_zero;
```

[0138] 如前所述,来自线性预测的特征提供了输入信号的电平独立分析,其改进了对背景噪声更新的判断,这在SNR范围10至20dB中特别有用,在该范围中,基于能量的SAD由于语音信号的正常动态范围而具有有限性能。

[0139] 背景接近度特征还改进了背景噪声估计,原因在于其可以用于初始化和正常操作二者。在初始化期间,它可以允许主要具有低频内容的(较低电平)背景噪声(这对于汽车噪声而言是常见的)。此外,这些特征可以用于防止使用低能量帧的噪声更新(该低能量帧与当前背景估计相比在频率特征上具有较大差异),这表明当前帧可能是低电平活跃内容,且更新可以防止对具有相似内容的未来帧的检测。

[0140] 图8-10示出了在10dB SNR汽车噪声的背景下,各个参数或度量针对语音的表现如何。在图8-10中,点“•”各自表示帧能量。对于图8和图9a-c,能量已被除以10,以与基于G_0_2和G_2_16的特征更具可比性。这些图对应于包括两个话语的音频信号,其中第一话语的近似位置在帧1310-1420中,且对于第二话语,在帧1500-1610中。

[0141] 图8示出了在汽车噪声情况下针对10dB SNR语音的帧能量(/10)(点,“•”)以及特征G_0_2(圆圈,“○”)和Gmax_0_2(加号,“+”)。注意,在汽车噪声期间,G_0_2是8,原因在于可以使用模型阶数为2的线性预测来建模的信号中存在某种相关性。在话语期间,特征Gmax_0_2变得超过1.5(在这种情况下),并在语音突发之后下降至0。在判断逻辑的特定实现中,G max_0_2需要低于0.1以允许使用该特征的噪声更新。

[0142] 图9a示出了帧能量(/10)(点,“•”)以及特征G_2_16(圆圈,“○”)、G1_2_16(十字,“x”)、G2_2_16(加号,“+”)。图9b示出了帧能量(/10)(点,“•”)以及特征G_2_16(圆圈,“○”)、Gd_2_16(十字,“x”)和Gad_2_16(加号,“+”)。图9c示出了帧能量(/10)(点,“•”)以及特征G_2_16(圆圈,“○”)和Gmax_2_16(加号,“+”)。图9a-c中示出的图还与汽车噪声情况下的10dB SNR语音相关。为易于查看每个参数,将特征示出在三个图中。注意,在汽车噪声(即,外部话语)期间,G_2_16(圆圈“○”)刚好大于1,指示对于这种类型的噪声,来自较高模型阶数的增益较低。在话语期间,特征Gmax_2_16(加号,图9c中的“+”)增加,且然后开始回落到0。在判断逻辑的特定实现中,特征Gmax_2_16也必须变得低于0.1以允许噪声更新。在

本特定音频信号样本中,不发生这种情况。

[0143] 图10示出了汽车噪声情况下针对10dB SNR语音的帧能量(点,“•”)(本次未除以10)和特征nonstaB(加号,“+”)。在仅具有噪声的段期间,特征nonstaB在范围0-10中,且对于话语,其变得更大(原因在于对于语音而言频率特性不同)。然而,应当注意,即使在话语期间,也存在特征nonstaB落入0-10的范围内的帧。对于这些帧,可能存在进行背景噪声更新并由此更好地跟踪背景噪声的可能性。

[0144] 本文公开的解决方案还涉及一种在硬件和/或软件中实现的背景噪声估计器。

[0145] 背景噪声估计器,图11a-11c

[0146] 在图11a中以一般方式示出了背景噪声估计器的示例性实施例。将背景噪声估计器称为被配置用于估计音频信号中的背景噪声的模块或实体,该音频信号包括语音和/或音乐。背景噪声估计器1100被配置为执行与上述例如参照图2和图7所描述的方法相对应的至少一种方法。背景噪声估计器1100关联于与前述方法实施例相同的技术特征、目的和优点。为了避免不必要的重复,将简要描述背景噪声估计器。

[0147] 可以如下实现和/或描述背景噪声估计器。

[0148] 背景噪声估计器1100被配置用于估计音频信号的背景噪声。背景噪声估计器1100包括处理电路或处理装置1101和通信接口1102。处理电路1101被配置为使背景噪声估计器1100基于以下各项获得(例如确定或计算)至少一个参数(例如NEW_POS_BG):第一线性预测增益,被计算为:针对音频信号段,来自0阶线性预测的残余信号与来自2阶线性预测的残余信号之间的商;以及,第二线性预测增益,被计算为:针对所述音频信号段,来自2阶线性预测的残余信号与来自16阶线性预测的残余信号之间的商。

[0149] 处理电路1101还被配置为使背景噪声估计器至少基于所获得的至少一个参数确定音频信号段是否包括暂停,即没有诸如语音和音乐的活跃内容。处理电路1101还被配置为使背景噪声估计器在音频信号段包括停顿时基于所述音频信号段更新背景噪声估计。

[0150] 通信接口1102,其也可以被表示为例如输入/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。例如,可以获得(例如,经由I/O接口从执行线性预测编码的音频信号编码器接收)与线性预测模型阶数0、2和16相关的残余信号。

[0151] 如图11b中所示,处理电路1101可以包括处理装置(例如处理器1103(例如CPU))和用于存储或保持指令的存储器1104。然后,存储器将包括例如具有计算机程序1105的形式的指令,所述指令当由处理装置1103执行时,使背景噪声估计器1100执行上述动作。

[0152] 在图11c中示出了处理电路1101的备选实现。这里的处理电路包括获得或确定单元或模块1106,被配置为使背景噪声估计器1100基于以下各项获得(例如确定或计算)至少一个参数(例如NEW_POS_BG):第一线性预测增益,被计算为:针对音频信号段,来自0阶线性预测的残余信号与来自2阶线性预测的残余信号之间的商;以及,第二线性预测增益,被计算为:针对所述音频信号段,来自2阶线性预测的残余信号与来自16阶线性预测的残余信号之间的商。处理电路还包括确定单元或模块1107,其被配置为使背景噪声估计器1100至少基于所获得的至少一个参数确定音频信号段是否包括暂停,即没有诸如语音和音乐的活跃内容;处理电路1101还包括更新单元1108,其被配置为使背景噪声估计器在音频信号段包括停顿时基于所述音频信号段更新背景噪声估计。

[0153] 处理电路1101可以包括更多单元,例如滤波单元或模块,其被配置为使背景噪声估计器对线性预测增益进行低通滤波,由此创建线性预测增益的一个或多个长期估计。例如低通滤波的动作可以通过其他方式执行,例如由确定单元或模块1107执行。

[0154] 上述背景噪声估计器的实施例可以被配置用于本文描述的不同方法实施例,例如对线性预测增益进行限制和低通滤波;确定线性预测增益和长期估计之间以及长期估计之间的差;和/或获得和使用谱接近度测量等。

[0155] 背景噪声估计器1100可以被假定为包括用于执行背景噪声估计的另外的功能(例如附录A中例示的功能)。

[0156] 图12示出了根据示例实施例的背景噪声估计器1200。背景估计器1200包括例如用于接收模型阶数0、2和16的残余能量的输入单元。背景估计器还包括处理器和存储器,所述存储器包含能够由所述处理器执行的指令,使得所述背景估计器操作用于:执行根据本文描述的实施例的方法。

[0157] 因此,如图13所示,背景估计器可以包括输入/输出单元1301、用于从模型阶数0、2和16的残余能量计算前两个特征集合的计算器1302、以及用于计算谱接近度特征的分析仪1303。

[0158] 如上所述的背景噪声估计器可以例如包括在VAD或SAD、编码器和/或解码器(即编解码器)中,和/或包括在设备(例如通信设备)中。通信设备可以是用户设备(UE),其形式为移动电话、视频相机、录音机、平板电脑、台式电脑、膝上型电脑、TV机顶盒或家庭服务器/家庭网关/家庭接入点/家庭路由器。在一些实施例中,通信设备可以是适于音频信号的编码和/或转码的通信网络设备。这种通信网络设备的示例是服务器,例如媒体服务器、应用服务器、网关和无线电基站。通信设备还可以适于布置在(即嵌入在)诸如轮船之类的船舰、无人飞机、飞机和诸如汽车、巴士或火车之类的道路车辆中。这种嵌入式设备通常属于车辆信息设备单元或车辆信息娱乐系统。

[0159] 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。

[0160] 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)。

[0161] 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。

[0162] 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。

[0163] 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。

[0164] 还应当理解,可以重用任意常用设备或者实现提出的技术的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。

[0165] 仅作为示例,提出上述实施例,并应当理解,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。

[0166] 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"。

[0167] 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。

[0168] 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,从而能够执行所建议的处理动作。

[0169] 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。

[0170] 除非明确地阐述,单数形式的元件的参考不意图表示“一个且仅一个”,而是“一个或多个”。上述优选元素实施例的元素的对于本领域的普通技术人员已知的所有结构和功能等同物明确通过引用并入本文,并旨在由本权利要求所涵盖。此外,设备或方法不必须解决本文公开的技术所要解决的每个问题,其用于被包含于此。

[0171] 在本文的一些实例中,省略众所周知的设备、电路和方法的详细说明,以免不必要的细节模糊所公开技术的说明。本文中列出所公开技术的原理、方面和实施例,以及其具体实例的所有陈述旨在包括其结构和功能等同物。此外,不考虑结构,希望这种等价形式既包括当前已知的等价形式,也包括未来发展的等价形式,例如执行相同功能的发展的单元。

[0172] 附录A

[0173] 下文中对附图的参考是参考图14和图15。

[0174] 图14是示出了根据本文提出的技术的用于背景噪声估计的方法的示例性实施例的流程图。意在由背景噪声估计器(其可以是SAD的一部分)来执行所述方法。背景噪声估计器和SAD还可以包括在音频编码器中,进而包括在无线设备或网络节点中。对于所描述的背景噪声估计器,不限制将噪声估计向下调节。对于每个帧,无论帧是背景还是活跃内容,如果新值低于它直接使用的当前值,则计算可能的新子带噪声估计,原因在于它最可能来自背景帧。以下的噪声估计逻辑是第二步骤,其中判断子带噪声估计是否可以增加并且如果可以增加的话可以增加多少,所述增加基于之前计算的可能的新子带噪声估计。基本上,该逻辑导致将当前帧判断为背景帧,并且如果其不确定的话,则可以允许比原始所估计的更

小的增加。

[0175] 图14所示的方法包括:当音频信号段的能量电平比高于(1402:1)长期最小能量电平lt_min的阈值大时,或者当音频信号段的能量电平比高于(1402:2)lt_min的阈值小,但是在音频信号段中没有检测到停顿(1404:1)时:

[0176] -当音频信号段被确定(1403:2)为包括音乐,并且当前背景噪声估计超过最小值(在图14中被表示为"T",并且在下面的代码中还被例示为例如2*E_MIN)(1405:1)时,减少(1406)当前背景噪声估计。

[0177] 通过执行上述操作,并向SAD提供背景噪声估计,使得SAD能够执行更充分的声音活跃性检测。此外,还能够从错误的背景噪声估计更新中恢复。

[0178] 在上述方法中使用的音频信号段的能量电平可以备选地称为例如当前帧能量(Etot),或者被称为信号段或帧的能量,其可以通过对当前信号段的子带能量进行求和来计算。

[0179] 在上述方法中使用的其它能量特征(即长期最小能量电平lt_min)是一种估计,其通过多个在先音频信号段或帧来确定。lt_min可以备选地表示为例如Etot_l_lp。推导lt_min的一个基本方式是使用若干过去帧的当前帧能量的历史的最小值。如果值被计算为:"当前帧能量-长期最小估计"低于阈值(被表示为例如THR1),则当前帧能量在此被视为接近长期最小能量,或者在长期最小能量附近。也就是说,当(Etot-lt_min)<THR1时,当前帧能量(Etot)可以被确定(1402)为在长期最小能量lt_min附近。取决于实现,当(Etot-lt_min)=THR1时的情况可以被称为判断1402:1或1402:2。图14中是序号1402:1指示当前帧能量不在lt_min附近的判断,而序号1402:2指示当前帧能量在lt_min附近的判断。图14中的关于形式XXX:Y的其它序号指示对应判断。将在下文进一步描述特征lt_min。

[0180] 当前背景噪声估计要超过的以便被减小的最小值可以被假设为零或小的正值。例如,如将在下面的代码中所解释的,背景估计的当前总能量(可以被表示为"totalNoise"并被确定为例如10*log10∑backr[i])需要超过最小值零以在后续讨论中减小。备选地或另外,可以将包含子带背景估计的矢量backr[i]中的每个条目与最小值E_MIN相比较,以执行减少。在下文的代码示例中,E_MIN是小的正值。

[0181] 应当注意,根据本文提出的解决方案的优选实施例,音频信号段的能量电平是否大于阈值(其高于lt_min)的判断仅基于从输入音频信号导出的信息,即它不基于来自声音活跃性检测器判断的反馈。

[0182] 可以基于一个或多个准则用不同的方式执行确定1404当前帧是否包括停顿。停顿准则也可以被称为停顿检测器。可以应用单个停顿检测器或不同停顿检测器的组合。在停顿检测器的组合的情况下,每个停顿检测器可以用于检测不同条件下的停顿。对当前帧可以包括停顿或不活跃的一个指示符是该帧的相关性特征较低,且多个在先帧也具有低相关性特征。如果当前能量接近长期最小能量,并且检测到停顿,则可以根据当前输入来更新背景噪声,如图14中所示。除了音频信号段的能量电平比高于lt_min的阈值小之外,可以在以下情况中视为检测到停顿:已经确定预定数量的连续在先音频信号段不包括活跃信号和/或音频信号的动态性超过阈值。这还示出在下文的代码示例中。

[0183] 背景噪声估计的减小(1406)使得能够处理背景噪声估计变得"太高"(即与真实背景噪声相关)的情况。这还可以被表达为背景噪声估计偏离了实际背景噪声。过高的背景噪

声估计可能导致SAD的不适当判断,其中,即使当前信号段包括活跃语音或音乐,其也被确定为是不活跃的。背景噪声估计变得太高的原因在于例如在音乐中的错误的或不期望的背景噪声更新,其中噪声估计将音乐误认为背景并且允许噪声估计增加。所公开的方法允许例如在输入信号的后续帧被确定为包括音乐时对错误更新的背景噪声估计进行调整。通过背景噪声估计的强制减小(其中噪声估计被缩减)来进行该调整,即使当前输入信号段能量高于例如子带中的当前背景噪声估计。应注意的是,上述用于背景噪声估计的逻辑用于控制背景子带能量的增加。在当前帧子带能量低于背景噪声估计时,始终允许降低子带能量。该功能没有在图14中明确地示出。这种下降通常具有针对步长的固定设置。然而,根据上述方法,背景噪声估计应该仅被允许与判断逻辑相关联地增加。当检测到停顿时,还可以使用能量和相关性特征来决定(1407)在进行实际背景噪声更新之前,背景估计增加的调整步长应该多大。

[0184] 如前所述,一些音乐段可能由于与噪声很像而难以从背景噪声中分离出来。因此,即使输入信号是活跃信号,噪声更新逻辑也可能意外地允许增加的子带能量估计。这会造成问题,因为噪声估计可能变得比它们所应该的高。

[0185] 在现有技术的背景噪声估计器中,仅在输入子带能量低于当前噪声估计时,子带能量估计才能减小。然而,由于一些音乐段可能由于很像噪声而难以从背景噪声中分离出来,发明人认识到需要针对音乐的恢复策略。在本文描述的实施例中,可以通过在输入信号回到类似音乐的特性时强制进行噪声估计减小来进行这种恢复。也就是说,当上文所述的能量和停顿逻辑防止(1402:1、1404:1)噪声估计增加时,测试(1403)输入是否被怀疑是音乐,如果是(1403:2),则逐帧将子带能量减小(1406)小的量,直到噪声估计达到最低电平(1405:2)。

[0186] 如上所述的背景估计器可以包括或者实现在VAD或SAD和/或编码器和/或解码器中,其中,编码器和/或解码器可以被实现在用户设备(例如移动电话、膝上型计算机、平板计算机等)中。背景估计器还可以包括在网络节点(例如媒体网关)中,例如作为编解码器的一部分。

[0187] 图15是示意性示出了根据示例性实施例的背景估计器的实现的框图。输入成帧块151首先将输入信号分成适当长度(例如5-30毫秒)的帧。对于每帧,特征提取器152从输入中计算至少以下特征:1)特征提取器在频域中分析帧,并且计算针对子带集合的能量。该子带是要用于背景估计的相同子带。2)特征提取器还分析时域中的帧,并计算相关性(表示为例如cor_est和/或lt_cor_est),其用于确定帧是否包括活跃内容。3)特征提取器还利用当前帧总能量(例如被表示为Etot)来更新当前和更早的输入帧的能量历史的特征,例如长期最小能量lt_min。相关性和能量特征随后被馈送给更新判断逻辑块153。

[0188] 这里,根据本文公开方案的判断逻辑被实现在更新判断逻辑块153中,其中,相关性和能量特征被用于判断当前帧能量是否接近长期最小能量;判断当前帧是否是停顿(非活跃信号)的一部分;以及判断当前帧是否是音乐的一部分。根据本文描述的实施例的解决方案涉及如何使用这些特征和判断以鲁棒的方式更新背景噪声估计。

[0189] 在下文中,将描述本文公开的方案的实施例的实施细节。下文的实施细节来自于基于G.718的编码器中的实施例。本实施例使用在WO2011/049514和WO2011/049515中描述的一些特征。

[0190] 在WO2011/049514中描述的修改的G.718中定义了以下特征：

[0191] Etot；当前输入帧的总能量

[0192] Etot_l 跟踪最小能量包络

[0193] Etot_l_lp；最小能量包络Etot_l的平滑版本

[0194] totalNoise；背景估计的当前总能量

[0195] bckr[i]；具有子带背景估计的矢量；

[0196] tmpN[i]；预计算的潜在新背景估计

[0197] aEn；使用多个特征的背景检测器(计数器)

[0198] harm_cor_cnt 对从具有相关性或谐波事件的最后帧开始的帧计数

[0199] act_pred 仅根据输入帧特征的活跃性预测

[0200] Cor[i]具有相关性估计的矢量,其中i＝0是当前帧的结束,

[0201] i＝1是当前帧的开始,i＝2是先前帧的结束

[0202] 在WO2011/049515中描述的修改的G.718中定义了以下特征：

[0203] Etot_h 跟踪最大能量包络

[0204] sign_dyn_lp；平滑后的输入信号动态特性

[0205] 在WO2011/049514中还定义了特征Etot_v_h,但是在本实施例中其被修改并且现在被实施为以下：

```
Etot_v = (float) fabs(*Etot_last - Etot);
if( Etot_v < 7.0f)        /*注意这里没有使用 VAD 标志或类似标志*/
{
        *Etot_v_h -= 0.01f;
        if (Etot_v > *Etot_v_h)
        {
```

[0206]

```
                if ((*Etot_v -*Etot_v_h) > 0.2f)
                {
                *Etot_v_h = *Etot_v_h + 0.2f;
                }
                else
                {
                        *Etot_v_h = Etot_v;          }}}
```

[0207] Etot_v测量帧之间的绝对能量变化,即,帧之间的瞬时能量变化的绝对值。在以上示例中,当最后帧能量和当前帧能量之间的差小于7个单位时,这两个帧之间的能量变化被确定为“低”。这被用作表明当前帧(和先前帧)可能是停顿的一部分(即,仅包括背景噪声)的指示符。然而,这种低变化还可以在例如突发语音的中间找到。变量Etot_last是先前帧的能量电平。

[0208] 以上在代码中描述的步骤可以被执行为图14中的流程图的“计算/更新相关性和能量”步骤的一部分,即,动作201的一部分。在WO2011/049514实现中,使用VAD标志来确定当前音频信号段是否包括背景噪声。发明人认识到依赖于反馈信息可能会有问题。在本文公开的方案中,确定是否更新背景噪声估计不依赖于VAD(或SAD)判断。

[0209] 此外,在本文公开的方案中,不是WO2011/049514实施的一部分的以下特征可以被计算/更新为相同步骤的一部分,即,在图14中示出的计算/更新相关性和能量步骤。这些特征还被判断逻辑用于确定是否更新背景估计。

[0210] 为了实现更准确的背景噪声估计,以下定义多个特征。例如,定义了新的关于相关性的特征cor_est和lt_cor_est。特征cor_est是当前帧中的相关性的估计,并且cor_est还用于产生lt_cor_est,lt_cor_est是对相关性的平滑后的长期估计。

[0211] cor_est＝(cor[0]+cor[1]+cor[2])/3.0f;

[0212] st->lt_cor_est＝0.01f*cor_est+0.99f*st->lt_cor_est;

[0213] 如上所述,cor[i]是包括相关性估计的矢量,cor[0]表示当前帧的结束,cor[1]表示当前帧的开始,并且cor[2]表示先前帧的结束。

[0214] 此外,计算新特征lt_tn_track,其给出背景估计接近当前帧能量的频繁程度的长期估计。在当前帧足够接近当前背景估计时,其被登记为用信号(1/0)通知是否接近背景的条件。该信号用于形成长期测量lt_tn_track。

[0215] st->lt_tn_track＝0.03f*(Etot-st->totalNoise<10)+0.97f*st->lt_tn_track;

[0216] 在该示例中,在当前帧能量接近背景噪声估计时,增加0.03,否则,仅剩余的项是先前值的0.97倍。在该示例中,“接近”被定义为当前帧能量Etot和背景噪声估计totalNoise之间的差小于10个单位。“接近”的其它定义也是可行的。

[0217] 此外,当前帧能量Etot和当前背景估计totalNoise之间的差被用于确定给出对该距离的长期估计的特征lt_tn_dist。创建类似的特征lt_Ellp_dist用于长期最小能量Etot_l_lp和当前帧能量Etot之间的距离。

[0218] st->lt_tn_dist＝0.03f*(Etot-st->totalNoise)+0.97f*st->lt_tn_dist;

[0219] st->lt_Ellp_dist＝0.03f*(Etot-st->Etot_l_lp)+0.97f*st->lt_Ellp_dist;

[0220] 上面介绍的特征harm_cor_cnt用于对从具有相关性或谐波事件的最后一帧开始(即,从满足与活跃性相关的某一准则的帧开始)的帧的数量进行计数。也就是说,当条件harm_cor_cnt＝＝0时,这意味着当前帧最可能是活跃帧,原因在于它示出了相关性或谐波事件。这用于形成对这种事件的发生频率的长期平滑估计lt_haco_ev。在此情况下,更新不对称,也就是说,在增加或降低估计的情况下使用不同时间常数,如下文所述。

```
if (st->harm_cor_cnt == 0)        /*当可能活跃时*/
{
    st->lt_haco_ev = 0.03f + 0.97f*st->lt_haco_ev;   /* 增加长期估
计*/
}
else
{
    st->lt_haco_ev = 0.99f*st->lt_haco_ev;      /*降低长期估计*/
}
```

[0221]

[0222] 上文引入的特征It_tn_track的低值指示针对一些帧,输入帧能量没有接近背景能量。这是由于在当前帧能量不接近背景能量估计的情况下,It_tn_track针对每个帧而降低。It_tn_track仅在当前帧能量接近背景能量估计时增加,如上所示。为了得到对该“不跟踪”(即,帧能量远离背景估计)持续了多久的更好估计,用于不存在跟踪的帧的数目的计数器low_tn_track_cnt被形成为:

```
if (st->lt_tn_track<0.05f)          /*当 lt_tn_track 为低时*/
{
    st->low_tn_track_cnt++;          /*将计数器加 1*/
}
else
{
    st->low_tn_track_cnt=0;          /*复位计数器*/
}
```

[0223]

[0224] 在以上示例中,“低”被定义为低于值0.05。这应该被视为示例性值,其可以被不同地选择。

[0225] 对于在图14中示出的步骤“形成停顿和音乐判断”,使用以下三个代码表达来形成停顿检测(还被表示为背景检测)。在其他实施例和实现中,针对停顿检测还可以添加其他准则。使用相关性和能量特征来在代码中形成实际音乐判断。

[0226] 1:bg_bgd=Etot<Etot_l_lp+0.6f*st->Etot_v_h;

[0227] 当Etot接近背景噪声估计时bg_bgd将变为“1”或“真”。bg_bgd用作针对其它背景检测器的掩码。也就是说,如果bg_bgd不为“真”,则不需要评估下面的背景检测器2和3。Etot_v_h是噪声变化估计,其可以备选地表示为N_{var}。使用Etot_v从(对数域中的)输入总能量得到Etot_v_h,其中,Etot_v测量帧之间的绝对能量变化。应注意,特征Etot_v_h被限制为将最大值仅增加小的常数值(例如,针对每帧,0.2)。Etot_l_lp是最小能量包络Etot_l的平滑版本。

[0228] 2:aE_bgd＝st->aEn＝＝0;

[0229] 当aEn为零时,aE_bgd变为“1”或“true”。aEn是当确定活跃信号存在于当前帧中时递增并当确定当前帧不包括活跃信号时递减的计数器。aEn可能不递增超过一定数量(例如6),并不减小到小于零。在多个(例如6个)连续帧之后,在没有活跃信号的情况下,aEn将等于零。

[0230] 3:

[0231] sd1_bgd＝(st->sign_dyn_lp>15)&&(Etot-st->Etot_l_lp)<st->Etot_v_h&&st->harm_cor_cnt>20;

[0232] 在以下三个不同情况是真的情况下,sd1_bgd将为“1”或“真”:信号动态性sign_dyn_lp高,在此示例中多于15;当前帧能量接近背景估计;以及:已通过了不具有相关性或谐波事件的某个数量的帧,在该示例中为20帧。

[0233] bg_bgd的功能是用于检测当前帧能量接近长期最小能量的标志。后两者(aE_bgd和sd1_bgd)表示不同条件下的停顿或背景检测。aE_bgd是这两者最常用的检测器,而sd1_bgd主要检测高SNR中的语音停顿。

[0234] 根据本文公开的技术的实施例的新判断逻辑被构建在下面的代码中。判断逻辑包括掩码条件bg_bgd以及两个停顿检测器aE_bgd和sd1_bgd。还可存在第三停顿检测器,其评估关于totalNoise跟踪最小能量估计的性能的长期统计量。在第一行为真的情况下评估的条件是关于步长updt_step应当多大的判断逻辑,且实际噪声更新是将值指派为“st->bckr[i]＝-”。应注意,tmpN[i]是之前计算的根据在WO2011/049514中描述的方案所计算的潜在新噪声电平。下文的判断逻辑遵循图14的部分1409,其被与下文的代码相关联地部分指示

```
if (bg_bgd && ( aE_bgd II sd1_bgd II st->lt_tn_track >0.90f ) ) /*如果 202:2
且 204:2)*/
        {
    if(   (st->act_pred < 0.85f II ( aE_bgd && st->lt_haco_ev < 0.05f ) )
    &&
        (st->lt_Ellp_dist < 10 II sd1_bgd ) && st->lt_tn_dist<40 &&
        ( (Etot - st->totalNoise ) < 15.0f II st->lt_haco_ev < 0.10f ) )
          /*207*/
              {
[0235]      st->first_noise_updt = 1;
        for( i=0; i< NB_BANDS; i++ )
        {
            st->bckr[i] = tmpN[i]          /*208*/
        }
              }
              else if (aE_bgd && st->lt_haco_ev < 0.15f)
              {
updt_step=0.1f;
if (st->act_pred > 0.85f )
```

```
        {
                updt_step=0.01f          /*207*/
        }
        if (updt_step > 0.0f)
        {
            st->first_noise_updt = 1;
          for[ i=0; i< NB_BANDS; i++ )
          {
          st->bckr[i] = st->bckr[i] + updt_step * (tmpN[i]-st->bckr[i]);
        /*208*/

                }}}
                else
              {
        (st->first_noise_updt) +=1;
[0236]          }
    }
    else
    {
        /*  如果在音乐中，进一步降低较低的 bckr */ /*如果 203:2 且 205:1*/
        If ( st->low_tn_track_cnt > 300 && st->lt_haco_ev > 0.9f && st->
    totalNoise > 0.0f)
        {
            For ( i=0; i< NB_BANDS; i++)
            {
                If (st->bckr[i] > 2 * E_MIN
                {
          St->bckr[i] = 0.98f * st->bckr[i];  /*206*/
                }
            }
```

```
            }
            Else
            {
[0237]
                    (st->first_noise_updt) += 1;
            }
        }
```

[0238] 以"/*如果在音乐中...*/开始的最后代码块中的代码段包含对背景估计的强制缩减,其在怀疑当前输入是音乐的情况下使用。这被判断为以下功能:与最小能量估计相比较长周期地较差跟踪背景噪声,以及,谐波或相关事件的频繁发生,以及,最后条件"totalNoise>0"是对背景估计的当前总能量大于零的检查,这意味着可以考虑减少背景估计。此外,确定是否"bckr[i]>2*E_MIN",其中E_MIN是小的正值。这是为了检查包括子带背景估计在内的矢量中的每个条目,从而条目需要超过E_MIN以被减小(在该示例中乘以0.98)。进行这些检查以避免将背景估计减小为太小的值。

[0239] 实施例改善了背景噪声估计,其使得SAD/VAD能够以更好的性能实现高效DTX方案,并且避免由于削波而引起的语音质量或音乐的劣化。

[0240] 通过从Etot_v_h移除在WO2011/049514中描述的判断反馈,可以更好地分离噪声估计和SAD。这是有益的,因为在如果/当SAD功能/调谐改变时,噪声估计不变。也就是说,背景噪声估计的确定变得独立于SAD的功能。此外,噪声估计逻辑的调谐也变得更简单,因为其不受当背景估计改变时来自于SAD的次生效应的影响。

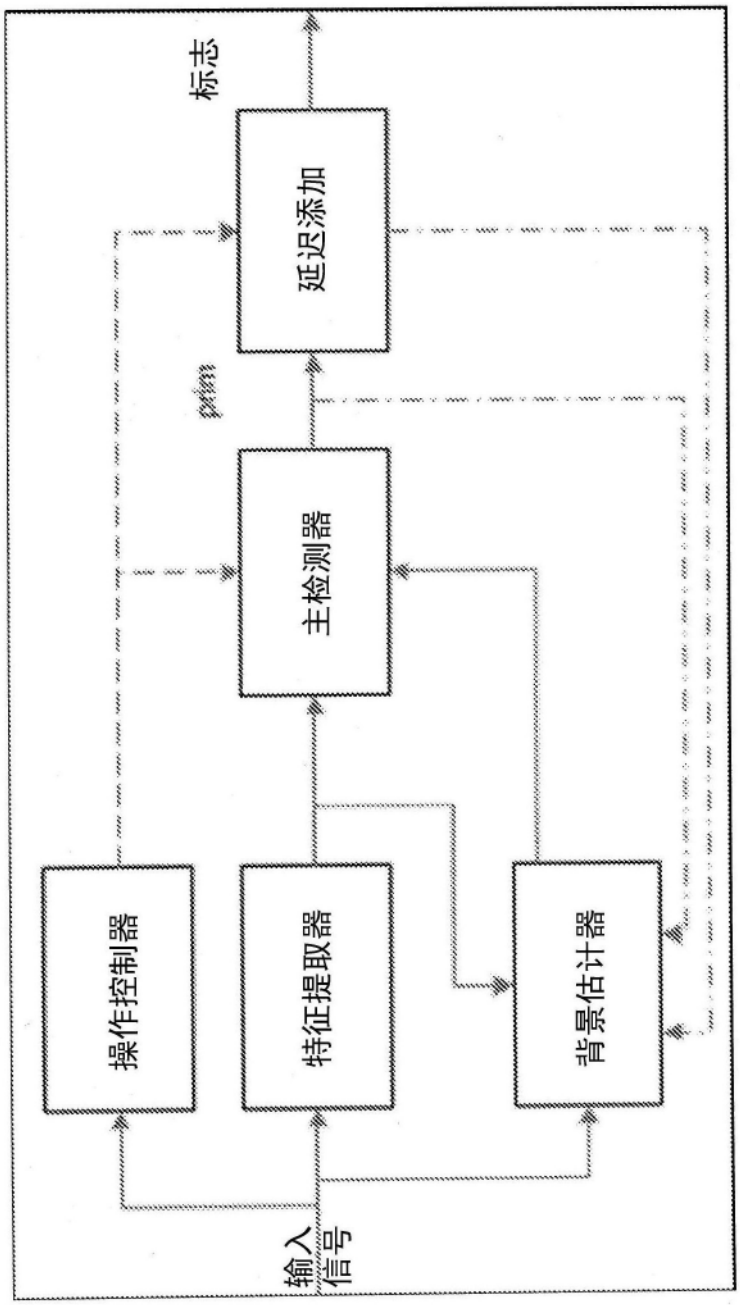


图1

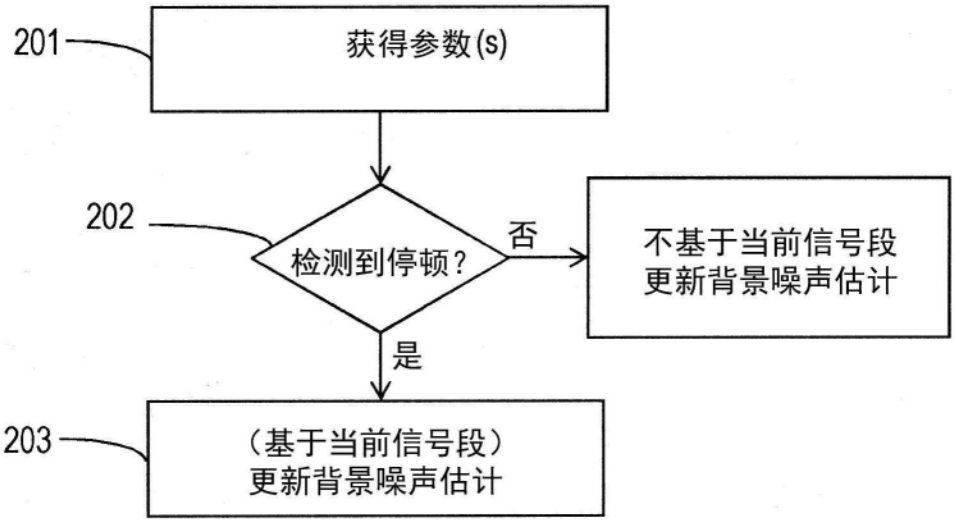

201
获得参数(s)
202
检测到停顿?
否
不基于当前信号段
更新背景噪声估计
是
203
(基于当前信号段)
更新背景噪声估计

图2

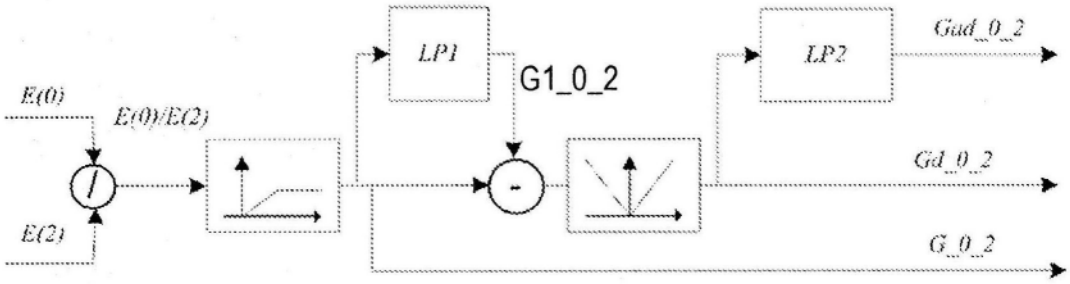

E(0)
E(0)/E(2)
E(2)
LP1
G1_0_2
LP2
Gad_0_2
Gd_0_2
G_0_2

图3

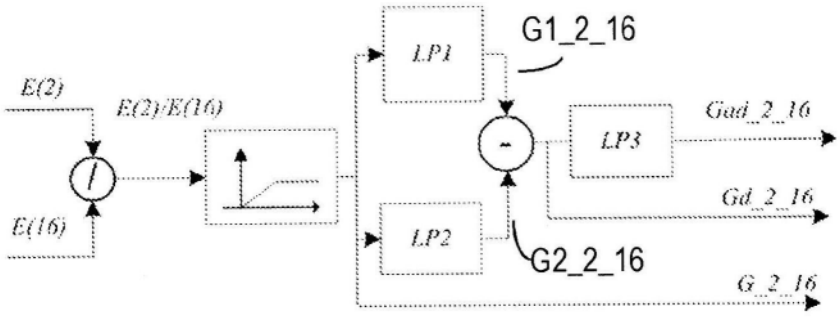

E(2)
E(2)/E(16)
E(16)
LP1
G1_2_16
LP3
Gad_2_16
Gd_2_16
LP2
G2_2_16
G_2_16

图4

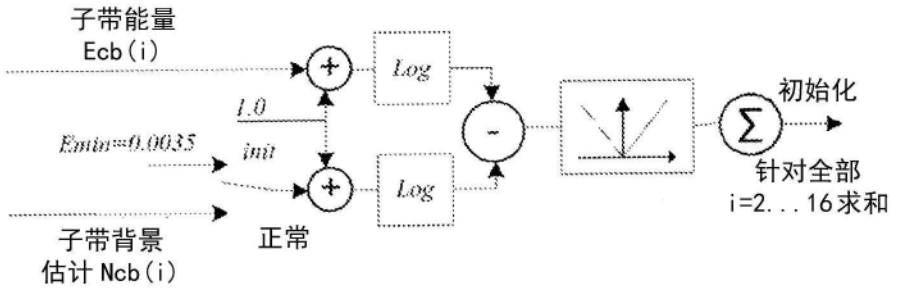

子带能量
Ecb(i)
Log
1.0
Emin=0.0035
init
Log
初始化
针对全部
i=2...16求和
子带背景
估计 Ncb(i)
正常

图5

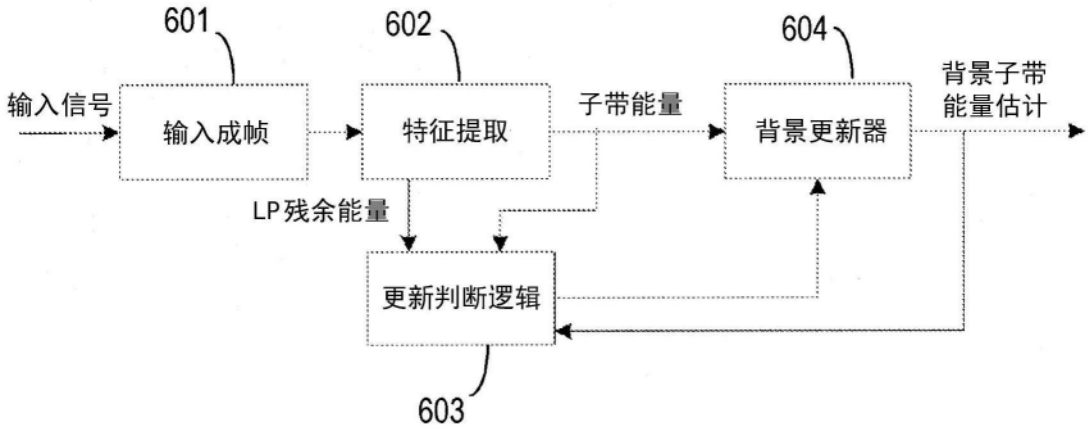

601
602
604
输入信号
输入成帧
特征提取
子带能量
背景更新器
背景子带
能量估计
LP残余能量
更新判断逻辑
603

图6

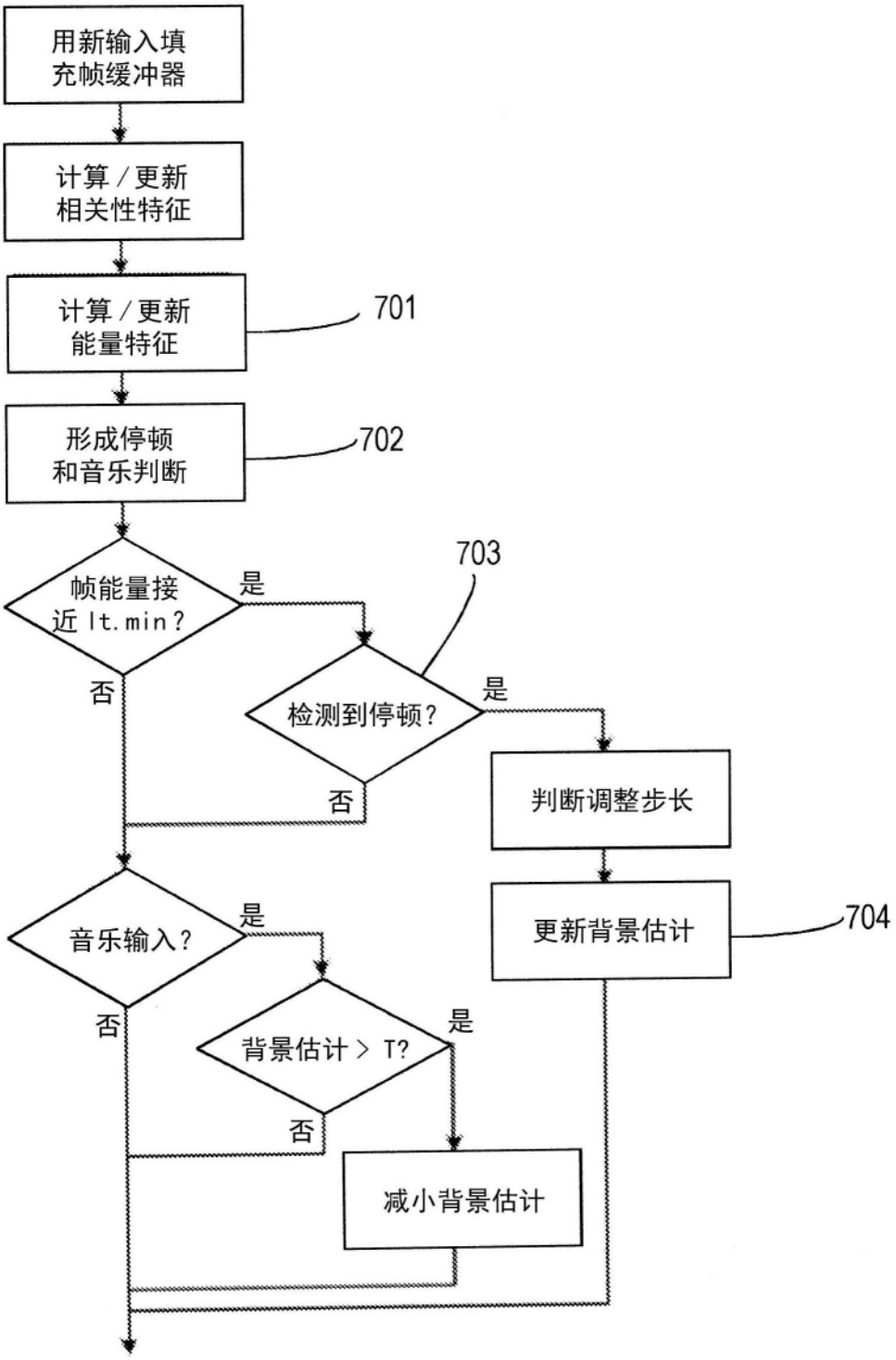

用新输入填充帧缓冲器
计算/更新相关性特征
计算/更新能量特征
701
形成停顿和音乐判断
702
帧能量接近 lt.min?
是
否
703
检测到停顿?
是
否
判断调整步长
更新背景估计
704
音乐输入?
是
否
背景估计 > T?
是
否
减小背景估计

图7

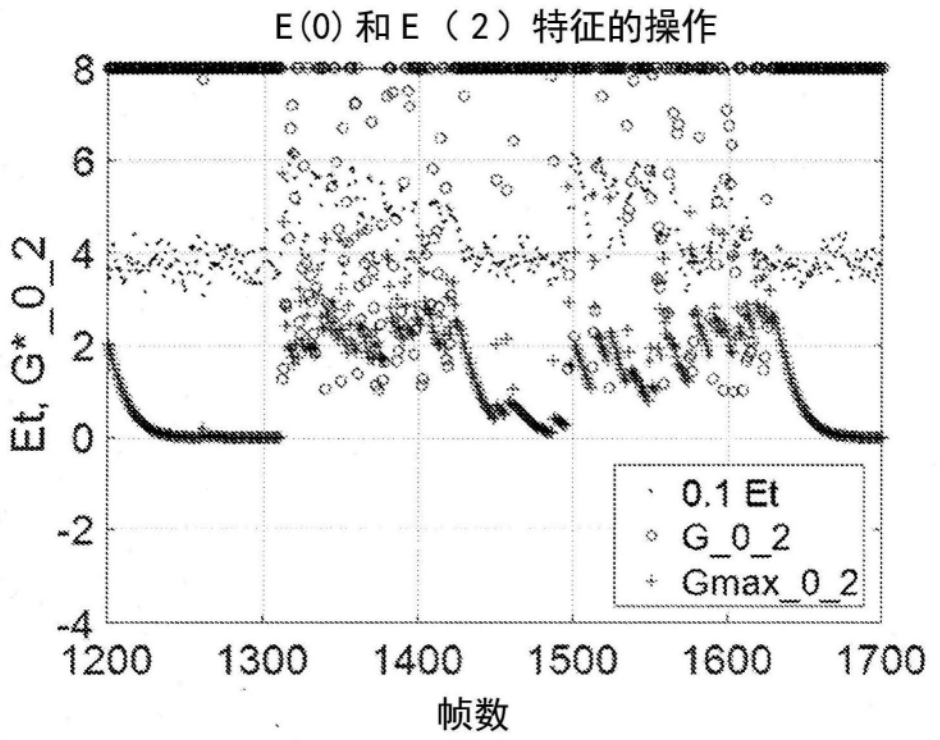

E(0)和E（2）特征的操作
Et, G*_0_2
帧数
0.1 Et
G_0_2
Gmax_0_2

图8

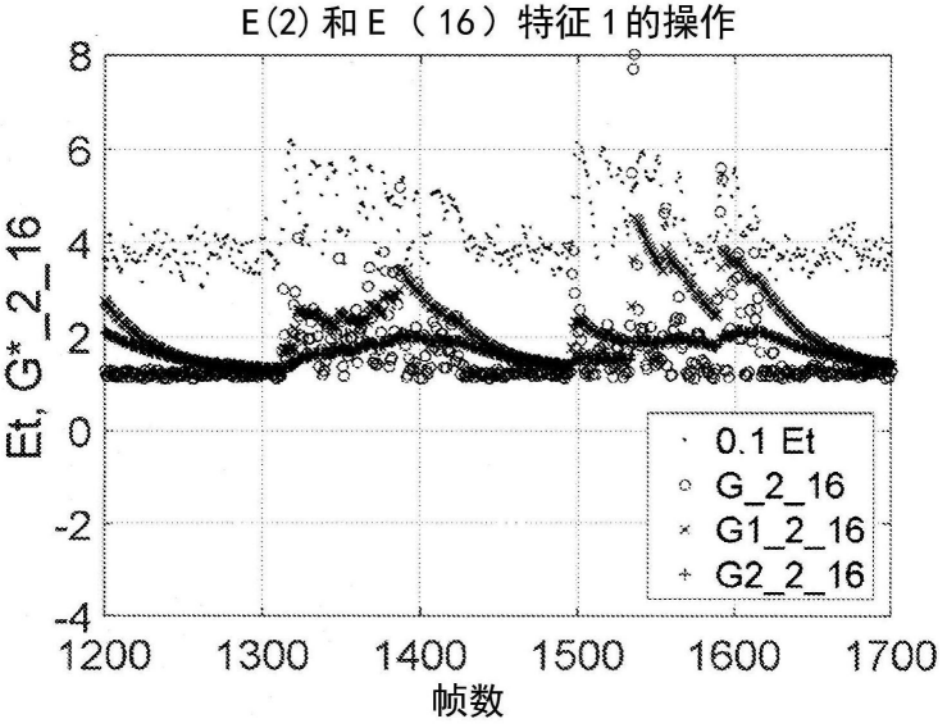

E(2)和E（16）特征1的操作
Et, G*_2_16
帧数
0.1 Et
G_2_16
G1_2_16
G2_2_16

图9a

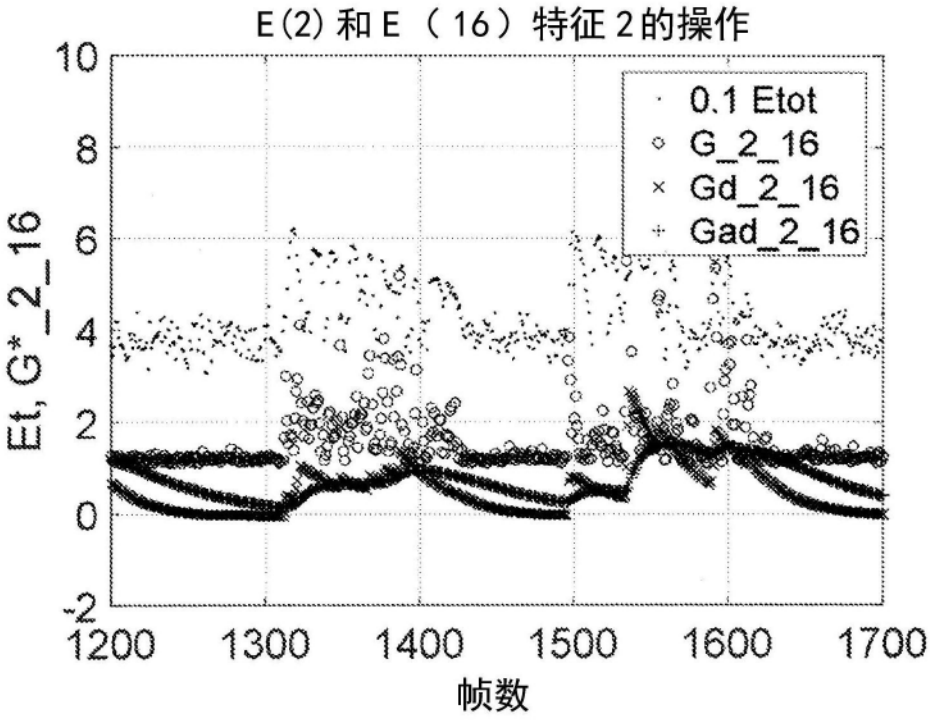

E(2) 和 E （ 16） 特征 2 的操作
0.1 Etot
G_2_16
Gd_2_16
Gad_2_16
Et, G*_2_16
帧数

图9b

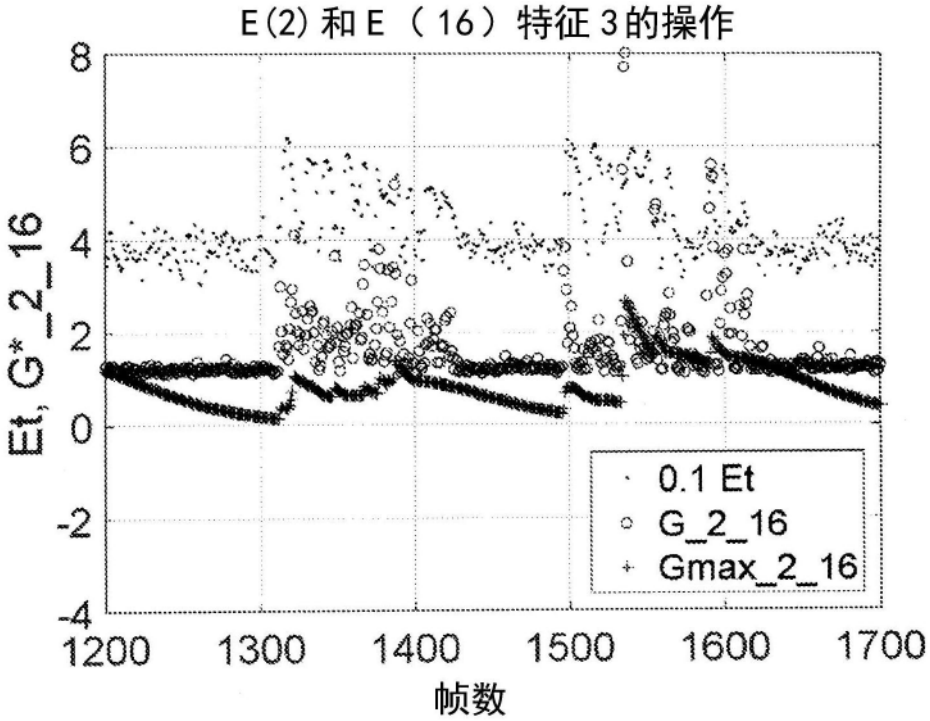

E(2) 和 E （ 16） 特征 3 的操作
0.1 Et
G_2_16
Gmax_2_16
Et, G*_2_16
帧数

图9c

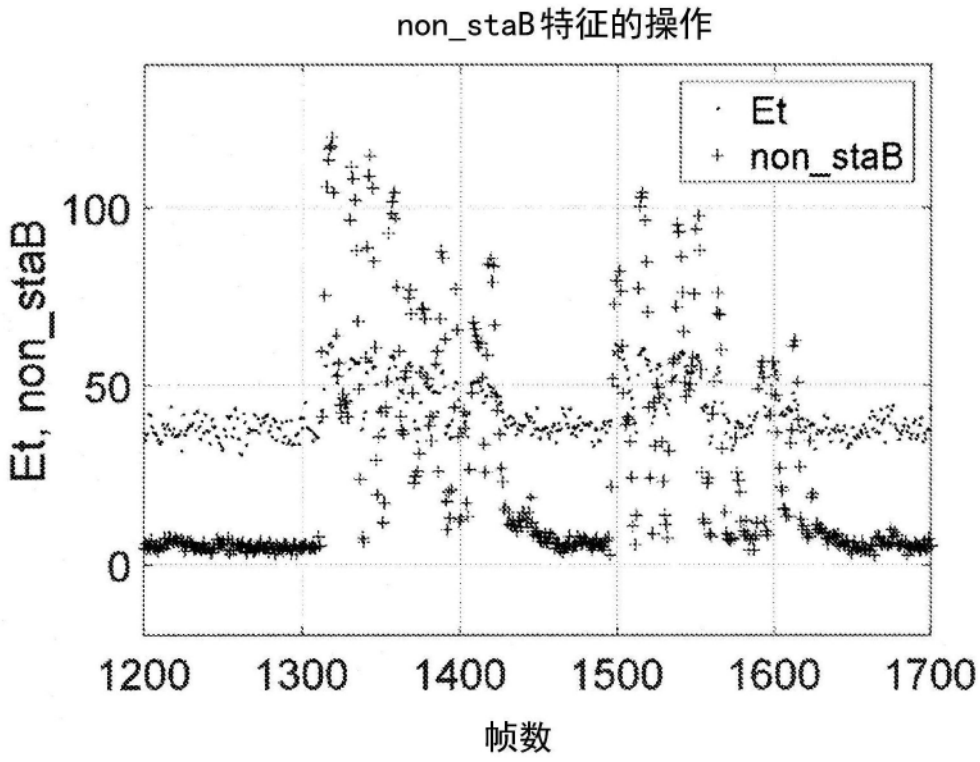

non_staB特征的操作
Et
non_staB
Et, non_staB
100
50
0
1200
1300
1400
1500
1600
1700
帧数

图10

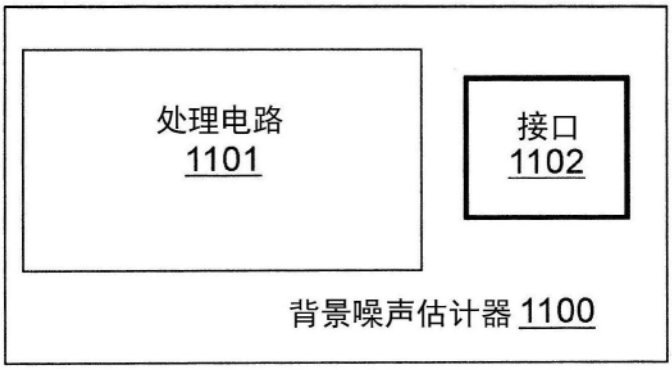

处理电路
1101
接口
1102
背景噪声估计器 1100

图11a

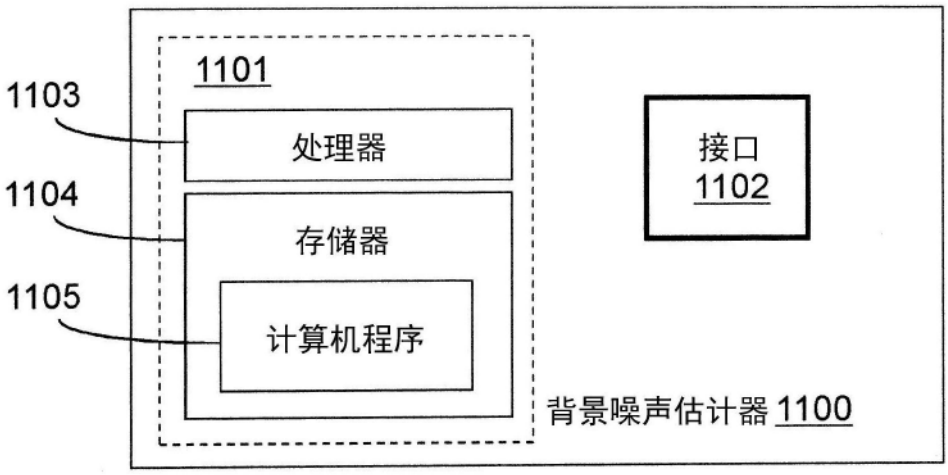

1101
1103
处理器
1104
存储器
1105
计算机程序
接口
1102
背景噪声估计器 1100

图11b

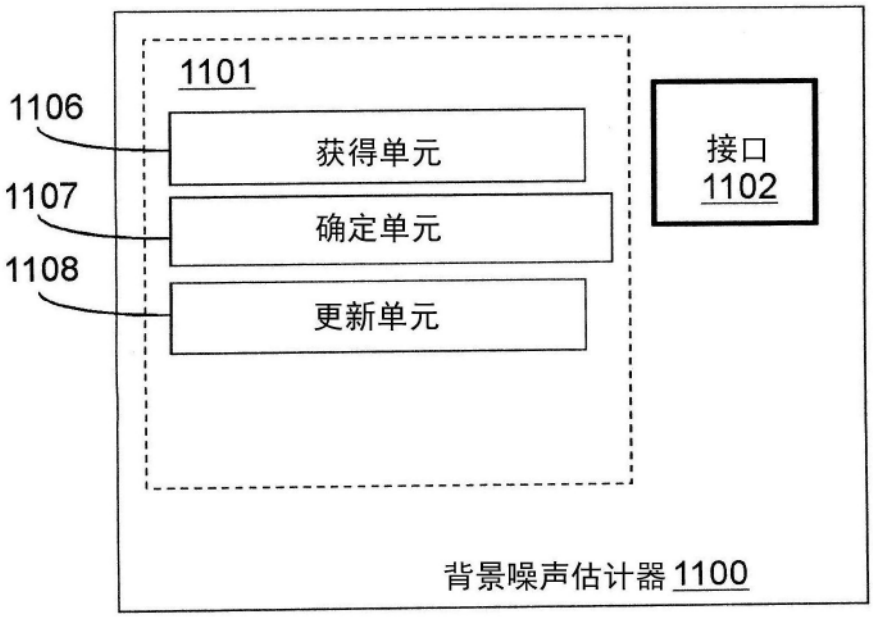

1101
1106
获得单元
1107
确定单元
1108
更新单元
接口
1102
背景噪声估计器 1100

图11c

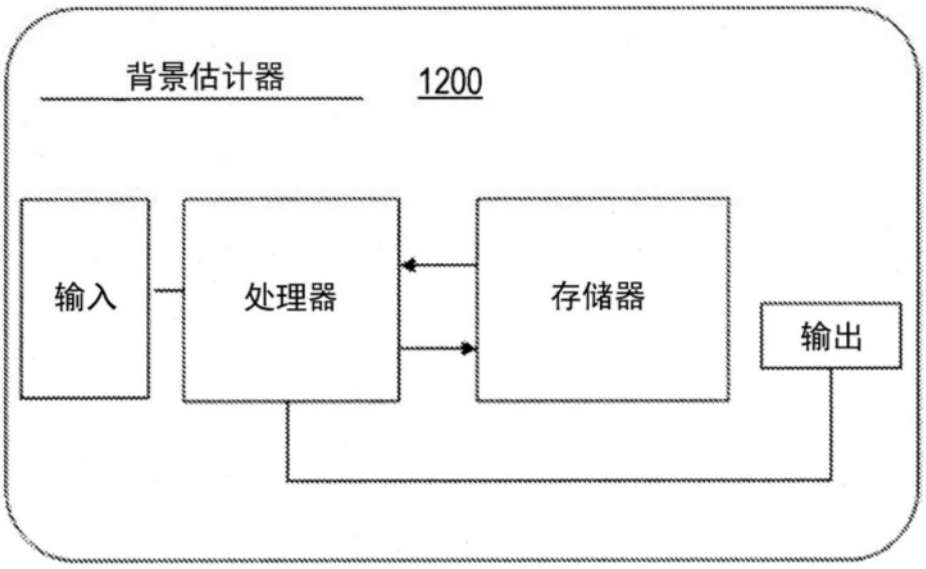

背景估计器
1200
输入
处理器
存储器
输出

图12

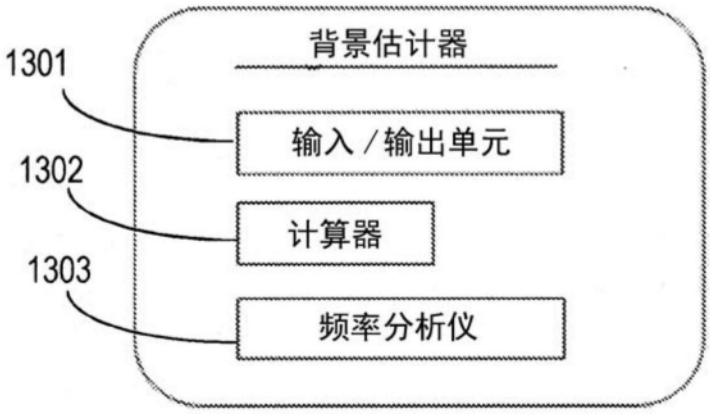

背景估计器
1301
输入/输出单元
1302
计算器
1303
频率分析仪

图13

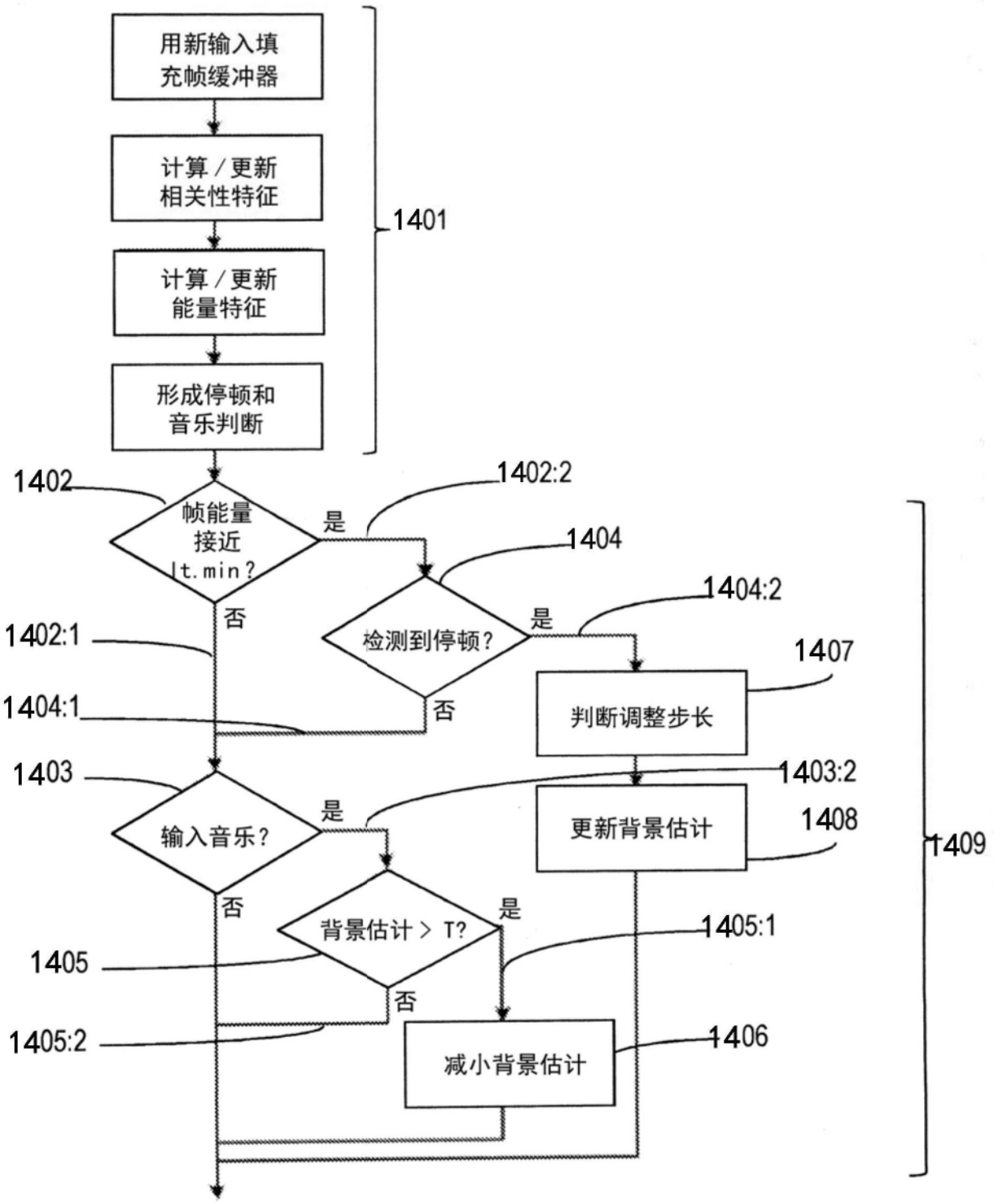

用新输入填充帧缓冲器
计算/更新相关性特征
计算/更新能量特征
形成停顿和音乐判断
1401
1402
帧能量接近 lt.min?
是
1402:2
否
1402:1
1404
检测到停顿?
是
1404:2
否
1404:1
1407
判断调整步长
1403
输入音乐?
是
1403:2
更新背景估计
1408
否
背景估计 > T?
是
1405:1
1405
否
1405:2
减小背景估计
1406
1409

图14

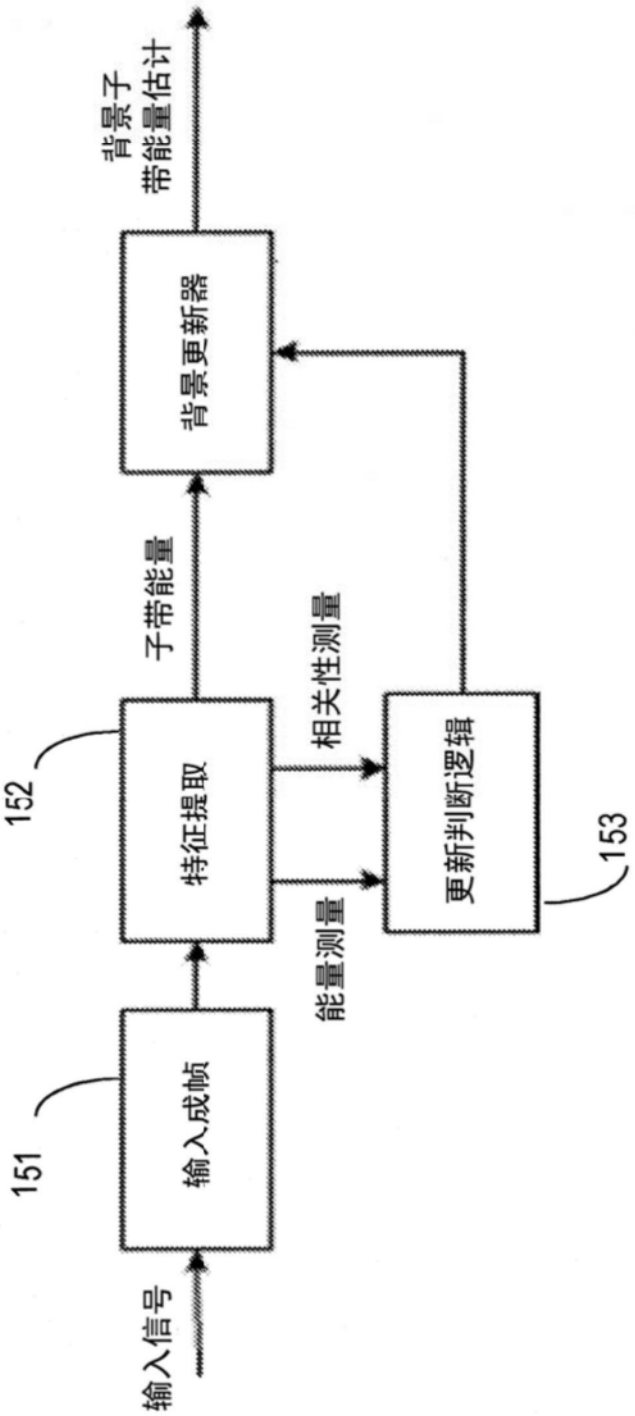

输入信号
151
输入成帧
152
特征提取
子带能量
背景更新器
背景子带能量估计
能量测量
相关性测量
更新判断逻辑
153

图15