



(12) 发明专利申请

(10) 申请公布号 CN 117392990 A

(43) 申请公布日 2024. 01. 12

(21) 申请号 202311306515.X

(51) Int. Cl.

(22) 申请日 2014.01.28

G10L 19/12 (2013.01)

(30) 优先权数据

G10L 19/002 (2013.01)

61/758,189 2013.01.29 US

G10L 19/028 (2013.01)

(62) 分案原申请数据

G10L 19/087 (2013.01)

201480019087.5 2014.01.28

G10L 19/26 (2013.01)

(71) 申请人 弗劳恩霍夫应用研究促进协会

地址 德国慕尼黑

(72) 发明人 纪尧姆·福奇斯

克里斯蒂安·赫尔姆里希

曼努埃尔·扬德尔

本杰明·苏伯特 横谷嘉一

(74) 专利代理机构 北京康信知识产权代理有限

责任公司 11240

专利代理师 刘彬

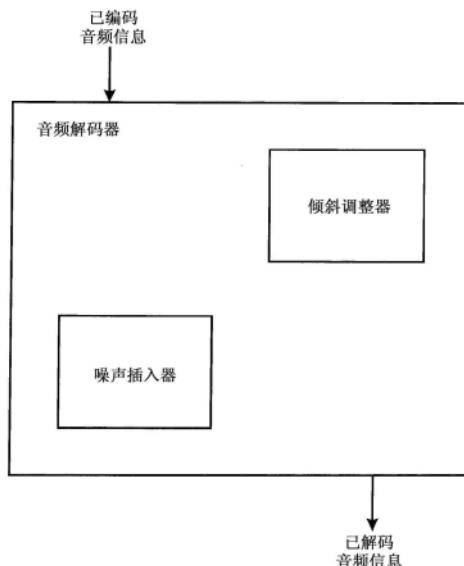
权利要求书3页 说明书11页 附图9页

(54) 发明名称

用于码激励线性预测类编码器的无边信息的噪声填充

(57) 摘要

本发明公开了用于码激励线性预测类编码器的无边信息的噪声填充。音频解码器包含：倾斜调整器，其被配置为使用当前帧的线性预测系数来调整噪声的倾斜以获得倾斜信息；以及噪声插入器，其被配置为根据由倾斜计算器获得的倾斜信息来将噪声添加至当前帧。根据本发明的另一音频解码器包含：噪声水平估计器，其被配置为使用至少一个先前帧的线性预测系数来估计当前帧的噪声水平以获得噪声水平信息；以及噪声插入器，其被配置为根据由噪声水平估计器提供的噪声水平信息来将噪声添加至当前帧。因此，可省略位流中的关于背景噪声的边信息。



1. 一种音频解码器,用于基于包括线性预测系数(LPC)的已编码音频信息来提供已解码音频信息,

所述音频解码器包含:

倾斜调整器,其被配置为使用当前帧的线性预测系数来调整噪声的倾斜以获得倾斜信息;以及

噪声插入器,其被配置为取决于由所述倾斜计算器获得的所述倾斜信息来将所述噪声添加至所述当前帧。

2. 根据权利要求1所述的音频解码器,其中,所述音频解码器包括用于判定所述当前帧的帧类型的帧类型判定器,所述帧类型判定器被配置为在所述当前帧的所述帧类型被检测为语音类型时,激活所述倾斜调整器来调整所述噪声的所述倾斜。

3. 根据权利要求1或2所述的音频解码器,其中,所述倾斜调整器被配置为使用所述当前帧的所述线性预测系数的一阶分析的结果来获得所述倾斜信息。

4. 根据权利要求3所述的音频解码器,其中,所述倾斜调整器被配置为使用所述当前帧的所述线性预测系数的增益 g 的计算作为所述一阶分析来获得所述倾斜信息。

5. 根据权利要求4所述的音频解码器,其中,所述倾斜调整器被配置为使用用于所述当前帧的直接形式滤波器 $x(n) - g \cdot x(n-1)$ 的传递函数的计算来获得所述倾斜信息。

6. 根据前述权利要求中任一项所述的音频解码器,其中,所述噪声插入器被配置为在将所述噪声添加至所述当前帧之前,将所述当前帧的所述倾斜信息应用于所述噪声以便调整所述噪声的所述倾斜。

7. 根据前述权利要求中任一项所述的音频解码器,其中,所述音频解码器还包含:

噪声水平估计器,其被配置为使用至少一个先前帧的线性预测系数来估计当前帧的噪声水平以获得噪声水平信息;以及

噪声插入器,其被配置为取决于由所述噪声水平估计器提供的所述噪声水平信息来将噪声添加至所述当前帧。

8. 一种音频解码器,用于基于包括线性预测系数(LPC)的已编码音频信息来提供已解码音频信息,

所述音频解码器包括:

噪声水平估计器,其被配置为使用至少一个先前帧的线性预测系数来估计当前帧的噪声水平以获得噪声水平信息;以及

噪声插入器,其被配置为取决于由所述噪声水平估计器提供的所述噪声水平信息来将噪声添加至所述当前帧。

9. 根据权利要求7或8所述的音频解码器,其中,所述音频解码器包括用于判定所述当前帧的帧类型的帧类型判定器,所述帧类型判定器被配置为识别所述当前帧的所述帧类型为语音还是一般音频,使得能取决于所述当前帧的所述帧类型来执行所述噪声水平估计。

10. 根据权利要求7至9中任一项所述的音频解码器,其中,所述音频解码器适于:计算表示所述当前帧的未频谱整形的激发的第一信息,计算关于所述当前帧的频谱缩放的第二信息,以及计算所述第一信息与所述第二信息的商来获得所述噪声水平信息。

11. 根据权利要求10所述的音频解码器,其中,所述音频解码器适于:在所述当前帧为语音类型的条件下,解码所述当前帧的激发信号,且从所述当前帧的时域表示来计算其均

方根 e_{rms} 来作为所述第一信息,以获得所述噪声水平信息。

12. 根据权利要求10或11所述的音频解码器,其中,所述音频解码器适于:在所述当前帧为语音类型的条件下,计算所述当前帧的LPC滤波器的传递函数的峰值水平 p 来作为第二信息,从而使用线性预测系数来获得所述噪声水平信息。

13. 根据权利要求11和12所述的音频解码器,其中,所述音频解码器适于:在所述当前帧为语音类型的条件下,通过计算所述均方根 e_{rms} 与所述峰值水平 p 的所述商来计算所述当前音频帧的频谱最小值 m_f ,以获得所述噪声水平信息。

14. 根据权利要求10至13所述的音频解码器,其中,所述音频解码器适于:如果所述当前帧为一般音频类型,则解码所述当前帧的未整形的MDCT激发,以及从所述当前帧的频谱域表示来计算其均方根 e_{rms} 来作为所述第一信息,以获得所述噪声水平信息。

15. 根据权利要求10至14中任一项所述的音频解码器,其中,所述音频解码器适于:无论帧类型如何,在所述噪声水平估计器中将从所述当前音频帧获得的所述商加入队列,所述噪声水平估计器包括用于从不同音频帧获得的两个或更多的商的噪声水平存储器。

16. 根据权利要求6或11所述的音频解码器,其中,所述噪声水平估计器适于:基于对不同音频帧的两个或更多的商的统计分析来估计所述噪声水平。

17. 根据前述权利要求中任一项所述的音频解码器,其中,所述音频解码器包括解码器核心,所述解码器核心被配置为使用所述当前帧的线性预测系数来解码所述当前帧的音频信息以获得已解码的核心编码器输出信号,并且其中,所述噪声插入器取决于在解码所述当前帧的所述音频信息时所使用的和/或在解码一个或多个先前帧的所述音频信息时所使用的线性预测系数来添加所述噪声。

18. 根据前述权利要求中任一项所述的音频解码器,其中,所述音频解码器包括去加重滤波器以将所述当前帧去加重,所述音频解码器适于在所述噪声插入器将所述噪声添加至所述当前帧之后对所述当前帧应用所述去加重滤波器。

19. 根据前述权利要求中任一项所述的音频解码器,其中,所述音频解码器包括噪声产生器,所述噪声产生器适于产生将由所述噪声插入器添加至所述当前帧的所述噪声。

20. 根据前述权利要求中任一项所述的音频解码器,其中,所述噪声产生器被配置为产生随机白噪声。

21. 根据前述权利要求中任一项所述的音频解码器,其中,所述噪声插入器被配置为在所述已编码音频信息的比特率小于每个样本1比特的条件下将所述噪声添加至所述当前帧。

22. 根据前述权利要求中任一项所述的音频解码器,其中,所述音频解码器被配置为使用基于编码器AMR-WB、G.718或LD-USAC (EVS) 中的一个或多个的编码器来解码所述已编码音频信息。

23. 一种用于基于包括线性预测系数(LPC)的已编码音频信息来提供已解码音频信息的方法,

所述方法包括:

使用当前帧的线性预测系数来调整噪声的倾斜以获得倾斜信息;以及

取决于所获得的倾斜信息来将所述噪声添加至所述当前帧。

24. 一种用于执行根据权利要求23所述的方法的计算机程序,其中,所述计算机程序在

计算机上运行。

25. 一种音频信号或储存有此音频信号的储存介质,所述音频信号已经用根据权利要求23所述的方法进行了处理。

26. 一种用于基于包括线性预测系数(LPC)的已编码音频信息来提供已解码音频信息的方法,

所述方法包括:

使用至少一个先前帧的线性预测系数来估计当前帧的噪声水平以获得噪声水平信息;
以及

取决于由所述噪声水平估计所提供的所述噪声水平信息来将噪声添加至所述当前帧。

27. 一种用于执行根据权利要求26所述的方法的计算机程序,其中,所述计算机程序在计算机上运行。

28. 一种音频信号或储存有此音频信号的储存介质,所述音频信号已经用根据权利要求26所述的方法进行了处理。

用于码激励线性预测类编码器的无边信息的噪声填充

[0001] 本申请是分案申请,其母案申请的申请号为201480019087.5,申请日为2014年1月28日,发明名称为“用于码激励线性预测类编码器的无边信息的噪声填充”。

技术领域

[0002] 本发明的实施方式涉及:用以基于包含线性预测系数(LPC)的已编码音频信息来提供已解码音频信息的音频解码器;用以基于包含线性预测系数(LPC)的已编码音频信息来提供已解码音频信息的方法;用以执行此方法的计算机程序,其中该计算机程序在计算机上运行;以及音频信号或储存有此音频信号的储存介质,该音频信号已经用此方法进行了处理。

背景技术

[0003] 当比特率降低至小于每个样本约0.5至1比特时,基于码激励线性预测(CELP)编码原理的低比特率数字语音(speech)编码器通常会遭受信号稀疏伪影,从而引起略为不自然的金属声。尤其当输入语音中具有背景中的环境噪声时,低速率(low-rate)伪影明显可听见:背景噪声在主动语音区段(active speech sections)期间将会衰减。本发明描述用于诸如AMR-WB[1]及G.718[4,7]的(A) CELP编码器的噪声插入方案,该方案与在诸如xHE-AAC[5,6]的基于变换的编码器中所使用的噪声填充技术类似,将随机噪声产生器的输出添加至已解码语音信号来重新建构背景噪声。

[0004] 国际公开案W0 2012/110476 A1展示出一种基于线性预测且使用频谱域噪声整形的编码概念。对音频输入信号的频谱分解(分解成包含连串频谱的频谱图)被用于以下两者:线性预测系数计算,以及用于基于线性预测系数的频域整形的输入。根据引用的文献,音频编码器包含线性预测分析器,其用以分析输入音频信号以便由此导出线性预测系数。音频编码器的频域整形器被配置为基于由线性预测分析器提供的线性预测系数频谱整形频谱图的一连串频谱的当前频谱。将已量化且已频谱整形的频谱连同在频谱整形时使用的线性预测系数一起插入至数据流中,使得在解码侧可执行去除整形(de-shaping)及去除量化(de-quantization)。也可存在时间噪声整形模块以执行时间噪声整形。

[0005] 鉴于现有技术,仍然需要改良的音频解码器、改良的方法、用以执行此方法的改良的计算机程序、以及改良的音频信号或储存有此音频信号的储存介质,该音频信号已经用此方法加以处理。更具体而言,需要找到改良在已编码位流中传递的音频信息的声音质量的解决方案。

发明内容

[0006] 在本发明的权利要求中和的实施方式的详细描述中的参考符号仅仅为了改善可读性而添加,绝不意味着是限制性的。

[0007] 本发明的目标是通过一种用以基于包含线性预测系数(LPC)的已编码音频信息来提供已解码音频信息的音频解码器来实现,该音频解码器包含:倾斜调整器(tilt

adjuster),其被配置为使用当前帧的线性预测系数来调整噪声的倾斜以获得倾斜信息;以及噪声插入器,其被配置为取决于由倾斜计算器获得的该倾斜信息来将该噪声添加至该当前帧。另外,本发明的目标通过一种用以基于包含线性预测系数(LPC)的已编码音频信息来提供已解码音频信息的方法来实现,该方法包含:使用当前帧的线性预测系数来调整噪声的倾斜以获得倾斜信息;以及取决于所获得的倾斜信息来将该噪声添加至该当前帧。

[0008] 作为第二种创造性解决方案,本发明建议一种用以基于包含线性预测系数(LPC)的已编码音频信息来提供已解码音频信息的音频解码器,该音频解码器包含:噪声水平估计器,其被配置为使用至少一个先前帧的线性预测系数来估计当前帧的噪声水平,以便获得噪声水平信息;以及噪声插入器,其被配置为取决于由该噪声水平估计器提供的该噪声水平信息来将该噪声添加至该当前帧。此外,本发明的目标是通过一种用以基于包含线性预测系数(LPC)的已编码音频信息来提供已解码音频信息的方法来解决,该方法包含:使用至少一个先前帧的线性预测系数来估计当前帧的噪声水平,以便获得噪声水平信息;以及取决于由该噪声水平估计器提供的噪声水平信息来将该噪声添加至该当前帧。另外,本发明的目标通过以下两者来解决:一种用以执行此方法的计算机程序,其中该计算机程序在计算机上运行;以及一种音频信号或储存有此音频信号的储存介质,该音频信号已经用此方法加以处理。

[0009] 所建议的解决方案避免了必须在CELP位流(bitstream,比特流)中提供边信息以便在噪声填充过程期间调整在解码器侧所提供的噪声。这意味着,可减小将要用位流传输的数据的量,而可仅仅基于当前或先前已解码的帧的线性预测系数来增加所插入噪声的质量。换言之,可省略关于噪声的边信息,该边信息将会增加将要用位流传递的数据的量。本发明允许提供低比特率数字编码器及方法,其与现有技术的解决方案相比而言可占用关于位流的更少的带宽并且提供质量提高的背景噪声。

[0010] 较佳的是,音频解码器包含用以判定当前帧的帧类型的帧类型判定器,该帧类型判定器被配置为在检测到当前帧的帧类型为语音类型时,启动倾斜调整器来调整噪声的倾斜。在一些实施方式中,帧类型判定器被配置为在帧经ACELP或CELP编码时,将该帧辨识为语音类型帧。根据当前帧的倾斜来对噪声加以整形可提供更自然的背景噪声且可减少与编码于位流中的所要信号的背景噪声有关的音频压缩的不良效应。因为这些不良的压缩效应及伪影相对于语音信息的背景噪声常常变得显著,所以可能有利的是:通过在将噪声添加至当前帧之前调整噪声的倾斜来增强将要添加至此类语音类型帧的噪声的质量。因此,噪声插入器可被配置为仅在当前帧为语音帧的情况下将噪声添加至当前帧,因为如果仅语音帧通过噪声填充来进行处理,可减少解码器侧的工作负载。

[0011] 在本发明的一较佳实施方式中,倾斜调整器被配置为使用对当前帧的线性预测系数的一阶分析(first-order analysis)的结果来获得倾斜信息。通过使用对线性预测系数此一阶分析,在位流中省略用以表征噪声的边信息成为可能。此外,对将要添加的噪声的调整可基于当前帧的线性预测系数,该等线性预测系数必须用位流以任何方式加以传递来允许对当前帧的音频信息的解码。这意味着在调整噪声的倾斜的过程中当前帧的线性预测系数被有利地再使用。另外,一阶分析相当简单,使得音频解码器的计算复杂性不会显著增加。

[0012] 在本发明的一些实施方式中,倾斜调整器被配置为使用对当前帧的线性预测系数

的增益 g 的计算作为该一阶分析来获得倾斜信息。更佳地,通过公式 $g = \sum [a_k \cdot a_{k+1}] / \sum [a_k \cdot a_k]$ 给出增益 g ,其中 a_k 为当前帧的LPC系数。在一些实施方式中,在该计算中使用两个或更多LPC系数 a_k 。更佳地,使用总共16个LPC系数,因此 $k=0 \cdots 15$ 。在本发明的实施方式中,位流可利用多于或少于16个LPC系数编码。因为当前帧的线性预测系数容易存在于位流中,所以可在不利用边信息的情况下获得倾斜信息,从而减小将要在位流中传递的数据的量。可仅仅通过使用对已编码音频信息加以解码所必需的线性预测系数来调整将要添加的噪声。

[0013] 更佳地,倾斜调整器可被配置为使用用于当前帧的直接形式滤波器 $x(n) - g \cdot x(n-1)$ 的传递函数的计算来获得倾斜信息。此种类型的计算相当容易且不需要解码器侧的高计算能力。如上文所展示,可易于根据当前帧的LPC系数计算出增益 g 。这允许在仅仅使用对已编码音频信息解码所必需的位流数据的情况下改善低比特率数字编码器的噪声质量。

[0014] 在本发明的一更佳实施方式中,噪声插入器被配置为在将噪声添加至当前帧之前,将当前帧的倾斜信息应用于噪声以便调整噪声的倾斜。若噪声插入器经相应地配置,则可提供简化的音频解码器。通过首先应用倾斜信息,随后将已调整的噪声添加至当前帧,可提供音频解码器的简单且有效的方法。

[0015] 在本发明的一实施方式中,音频解码器另外包含:噪声水平估计器,其被配置为使用至少一个先前帧的线性预测系数来估计当前帧的噪声水平以获得噪声水平信息;以及噪声插入器,其被配置为取决于由该噪声水平估计器提供的该噪声水平信息来将噪声添加至该当前帧。由此,因为可根据可能存在于当前帧中的噪声水平来调整将要添加至当前帧的噪声,所以可增强背景噪声的质量且因此增强整个音频传输的质量。例如,若因为根据先前帧估计了高噪声水平,所以预计在当前帧中为高噪声水平,则噪声插入器可被配置为在将噪声添加至当前帧之前增加将要添加至当前帧的噪声的水平。因此,将要添加的噪声可被调整成与当前帧中的预计噪声水平相比而言既不会太安静也不会太大声。此外,此调整并非基于位流中的专用边信息,而是仅仅使用在位流中传递的必要数据的信息,在此情况下为至少一个先前帧的线性预测系数,该线性预测系数亦提供关于先前帧中的噪声水平的信息。因此,更佳的是,使用 g 导出的倾斜对将要添加至当前帧的噪声加以整形且考虑到噪声水平估计来缩放(scale)该噪声。更佳的是,在当前帧为语音类型时,调整将要添加至当前帧的噪声的倾斜及噪声水平。在一些实施方式中,在当前帧为例如TCX类型或DTX类型的一般音频类型时,也调整将要添加至当前帧的倾斜和/或噪声水平。

[0016] 更佳地,音频解码器包含用以判定当前帧的帧类型的帧类型判定器,该帧类型判定器被配置为识别当前帧的帧类型为语音还是一般音频,因此可取决于当前帧的帧类型来执行噪声水平估计。例如,帧类型判定器可被配置为检测当前帧为CELP或ACELP帧(其是语音帧类型),还是TCX/MDCT或DTX帧(其是一般音频帧类型)。因为这些编码格式遵循不同原理,所以需要在执行噪声水平估计之前判定帧类型,以使得可取决于帧类型来选择适合的计算。

[0017] 在本发明的一些实施方式中,音频解码器适于:计算表示当前帧的未频谱整形的激发(excitation,激励)的第一信息,以及计算关于当前帧的频谱缩放的第二信息,以便计算第一信息和第二信息的商(quotient)来获得噪声水平信息。由此,可在不利用任何边信息的情况下获得噪声水平信息。因此,可保持编码器的比特率较低。

[0018] 较佳地, 音频解码器适于: 在当前帧为语音类型的条件下, 解码当前帧的激发信号, 且根据当前帧的时域表示来计算其均方根 e_{rms} 来作为第一信息, 以便获得噪声水平信息。对此实施方式较佳的是, 音频解码器适于在当前帧为CELP或ACELP类型的情况下相应地执行。将已频谱整平的激发信号(在感知域中)从位流解码且用来更新噪声水平估计。在读取位流之后计算当前帧的激发信号的均方根 e_{rms} 。此种类型的计算可不需要高计算能力, 因此甚至可由具有较低计算能力的音频解码器执行。

[0019] 在一较佳实施方式中, 音频解码器适于: 在当前帧为语音类型的条件下, 计算当前帧的LPC滤波器的传递函数的峰值水平 p 来作为第二信息, 从而使用线性预测系数来获得噪声水平信息。此外, 较佳的是, 当前帧为CELP或ACELP类型。计算峰值水平 p 的成本相当低, 且通过再使用当前帧的线性预测系数(也用来解码该帧中所含的音频信息), 可省略边信息, 且仍可增强背景噪声而不增加位流的数据速率。

[0020] 在本发明的一较佳实施方式中, 音频解码器适于: 在当前帧为语音类型的条件下, 通过计算均方根 e_{rms} 与峰值水平 p 的商来计算当前音频帧的频谱最小值 m_f , 以便获得噪声水平信息。此计算相当简单且可提供可用于估计在多个音频帧的范围内的噪声水平的数值。因此, 可使用一系列当前音频帧的频谱最小值 m_f 来估计在该等一系列音频帧所涵盖的时段期间的噪声水平。这可允许在保持复杂性相当低的同时获得对当前帧的噪声水平的良好估计。较佳地使用公式 $p = \sum |a_k|$ 来计算峰值水平 p , 其中 a_k 为线性预测系数, 较佳地, $k = 0 \cdots 15$ 。因此, 若帧包含16个线性预测系数, 则在一些实施方式中可通过对较佳为16个的 a_k 的振幅求和来计算 p 。

[0021] 较佳地, 音频解码器适于: 在当前帧为一般音频类型的情况下, 解码当前帧的未整形的MDCT激发, 且根据当前帧的频谱域表示来计算其均方根 e_{rms} 以便获得噪声水平信息来作为第一信息。每当当前帧并非语音帧, 而是一般音频帧时, 此系本发明的较佳实施方式。在MDCT或DTX帧中的频谱域表示很大程度上等效于在例如CELP或(A) CELP帧的语音帧中的时域表示。差别在于, MDCT未考虑帕塞瓦尔定理(Parseval's theorem)。因此, 较佳地, 计算一般音频帧的均方根 e_{rms} 的方式类似于计算语音帧的均方根 e_{rms} 的方式。然后, 较佳地, 如WO 2012/110476 A1中所述, 例如使用MDCT功率谱来计算一般音频帧的LPC系数等效物(LPC coefficients equivalents), 该MDCT功率谱指代巴克尺度(bark scale)上的MDCT值的平方。在替代实施方式中, MDCT功率谱的频带可具有恒定的宽度, 因此该功率谱的尺度对应于线性尺度(linear scale, 线性标尺)。在此线性尺度的情况下, 计算出的LPC系数等效物类似于例如针对ACELP或CELP帧所计算出的相同帧的时域表示中的LPC系数。另外, 较佳的是, 若当前帧为一般音频类型, 则计算如WO 2012/110476 A1中所述根据MDCT帧所计算出的当前帧的LPC滤波器的传递函数的峰值水平 p 来作为第二信息, 从而在当前帧为一般音频类型的条件下使用线性预测系数来获得噪声水平信息。然后, 若当前帧为一般音频类型, 则较佳地通过计算均方根 e_{rms} 和峰值水平 p 的商来计算当前音频帧的频谱最小值, 以便在当前帧为一般音频类型的条件下获得噪声水平信息。因此, 无论当前帧为语音类型还是一般音频类型, 均可获得描述当前帧的频谱最小值 m_f 的商。

[0022] 在一较佳实施方式中, 音频解码器适于: 无论帧类型如何, 在噪声水平估计器中将从当前音频帧获得的商加入队列, 该噪声水平估计器包含用于从不同音频帧获得的两个或更多商的噪声水平存储器。例如在应用低延迟统一语音及音频解码(LD-USAC、EVS)时, 如果

音频解码器适于在语音帧的解码与一般音频帧的解码之间切换,这会是有益的。由此,无论帧类型如何,均可获得多个帧的平均噪声水平。较佳地,噪声水平储存器可保存从十个或更多先前音频帧获得的十个或更多的商。例如,噪声水平储存器可含有用于30个帧的商的空间。因此,可针对在当前帧之前的扩展时间计算出噪声水平。在一些实施方式中,仅在检测到当前帧为语音类型时,可在噪声水平估计器中将商加入队列。在其他实施方式中,仅在检测到当前帧为一般音频类型时,可在噪声水平估计器中将商加入队列。

[0023] 较佳的是,噪声水平估计器适于基于不同音频帧的两个或更多商的统计分析来估计噪声水平。在本发明的一实施方式中,音频解码器适于使用基于最小均方误差的噪声功率谱密度追踪来对该等商进行统计分析。在Hendriks、Heusdens以及Jensen的公开案[2]中描述了此追踪。如果应该应用根据[2]的方法,则音频解码器适于在统计分析时使用轨迹值的平方根,就像在本例中一样直接搜寻振幅谱。在本发明的另一实施方式中,使用根据[3]已知的最小值统计数据来分析不同音频帧的两个或更多商。

[0024] 在一较佳实施方式中,音频解码器包含解码器核心,解码器核心被配置为使用当前帧的线性预测系数来解码当前帧的音频信息以获得已解码的核心编码器输出信号,且噪声插入器取决于在解码当前帧的音频信息时所使用的和/或在解码一个或多个先前帧的音频信息时所使用的线性预测系数来添加噪声。因此,噪声插入器利用用来解码当前帧的音频信息的相同线性预测系数。可省略用来指示噪声插入器的边信息。

[0025] 较佳地,音频解码器包含用以将当前帧去加重的去加重滤波器(de-emphasis filter),该音频解码器适于在噪声插入器将噪声添加至当前帧之后对当前帧应用去加重滤波器。由于去加重是提升低频的一阶IIR,所以这允许对所添加噪声的低复杂性、陡峭IIR高通滤波,从而避免在低频处的可听见的噪声伪影。

[0026] 较佳地,音频解码器包含噪声产生器,该噪声产生器适于产生将由噪声插入器添加至当前帧的噪声。使音频解码器包括噪声产生器可提供更方便的音频解码器,因为不需要外部噪声产生器。在替代方案中,噪声可由外部噪声产生器供应,外部噪声产生器可经由接口连接至音频解码器。例如,取决于在当前帧中将要增强的背景噪声,可应用特殊类型的噪声产生器。

[0027] 较佳地,噪声产生器被配置为产生随机白噪声。此噪声与常见的背景噪声充分相似,且此噪声产生器可易于提供。

[0028] 在本发明的一较佳实施方式中,噪声插入器被配置为在已编码音频信息的比特率小于每个样本1个比特的条件下将噪声添加至当前帧。较佳地,已编码音频信息的比特率小于每个样本0.8比特。甚至更佳的是,噪声插入器被配置为在已编码音频信息的比特率小于每个样本0.5比特的条件下将噪声添加至当前帧。

[0029] 在一较佳实施方式中,音频解码器被配置为使用基于编码器AMR-WB、G.718或LD-USAC (EVS) 中的一个或多个的编码器来解码已编码音频信息。这些编码器是熟知的且分布广泛的(A) CELP编码器,在这些编码器中额外使用这样的噪声填充方法会是极为有利的。

附图说明

[0030] 以下关于附图来描述本发明的实施方式。

[0031] 图1示出了根据本发明的音频解码器的第一实施方式;

[0032] 图2示出了根据本发明的用于执行音频解码的第一种方法,该方法可由根据图1的音频解码器执行;

[0033] 图3示出了根据本发明的音频解码器的第二实施方式;

[0034] 图4示出了根据本发明的用于执行音频解码的第二种方法,该方法可由根据图3的音频解码器执行;

[0035] 图5示出了根据本发明的音频解码器的第三实施方式;

[0036] 图6示出了根据本发明的用于执行音频解码的第三种方法,该方法可由根据图5的音频解码器执行;

[0037] 图7示出了用于计算用于噪声水平估计的频谱最小值 m_f 的方法的例示;

[0038] 图8示出了例示了从LPC系数导出的倾斜的图;以及

[0039] 图9示出了例示了如何根据MDCT功率谱确定LPC滤波器等效物的图。

具体实施方式

[0040] 关于图1至图9来详细描述本发明。本发明绝不意味着限于所示出及描述的实施方式。

[0041] 图1示出了根据本发明的音频解码器的第一实施方式。音频解码器适于基于已编码音频信息来提供已解码音频信息。音频解码器被配置为使用可基于AMR-WB、G.718及LD-USAC (EVS) 的编码器来解码已编码音频信息。已编码音频信息包含可分别表示为系数 a_k 的线性预测系数(LPC)。音频解码器包含:倾斜调整器,其被配置为使用当前帧的线性预测系数来调整噪声的倾斜以获得倾斜信息;以及噪声插入器,其被配置为取决于通过倾斜计算器获得的倾斜信息来将噪声添加至当前帧。噪声插入器被配置为在已编码音频信息的比特率小于每个样本1个比特的条件下将噪声添加至当前帧。另外,噪声插入器可被配置为在当前帧为语音帧的条件下将噪声添加至当前帧。因此,可将噪声添加至当前帧以便改善已解码音频信息的总体声音质量,该质量可能因编码伪影而受损,尤其就语音信息的背景噪声而言。当考虑到当前音频帧的倾斜来调整噪声的倾斜时,可在不取决于位流中的边信息的情况下改善总体声音质量。因此,可减小将要用位流传递的数据的量。

[0042] 图2示出了根据本发明的用于执行音频解码的第一种方法,该方法可由根据图1的音频解码器执行。连同方法特征一起描述了图1中所描绘的音频解码器的技术细节。音频解码器适于读取已编码音频信息的位流。音频解码器包含用于判定当前帧的帧类型的帧类型判定器,该帧类型判定器被配置为在检测到当前帧的帧类型为语音类型时,激活倾斜调整器来调整噪声的倾斜。因此,音频解码器通过应用帧类型判定器来判定当前音频帧的帧类型。若当前帧为ACELP帧,则帧类型判定器激活倾斜调整器。倾斜调整器被配置为使用对当前帧的线性预测系数的一阶分析的结果来获得倾斜信息。更具体而言,倾斜调整器使用公式 $g = \sum [a_k \cdot a_{k+1}] / \sum [a_k \cdot a_k]$ 作为一阶分析来计算增益 g ,其中 a_k 为当前帧的LPC系数。图8示出了例示了从LPC系数导出的倾斜的图。图8示出了单词「see」的两个帧。对于具有大量高频的字母「s」,倾斜向上。对于具有大量低频的字母「ee」,倾斜向下。图8所示的频谱倾斜是直接形式滤波器 $x(n) - g \cdot x(n-1)$ 的传递函数,其中 g 是如上文所述地定义。因此,倾斜调整器利用在位流中所提供的且用来解码已编码音频信息的LPC系数。因此可省略边信息,从而可减小将要用位流传递的数据的量。另外,倾斜调整器被配置为使用直接形式滤波器 $x(n) -$

$g \cdot x(n-1)$ 的传递函数的计算来获得倾斜信息。因此,倾斜调整器通过使用先前计算出的增益 g 计算出直接形式滤波器 $x(n) - g \cdot x(n-1)$ 的传递函数来计算当前帧中的音频信息的倾斜。在获得倾斜信息之后,倾斜调整器取决于当前帧的倾斜信息来调整将要添加至当前帧的噪声的倾斜。在此之后,将已调整的噪声添加至当前帧。另外,图2中未示出,音频解码器包含用于将当前帧去加重的去加重滤波器,音频解码器适于在噪声插入器将噪声添加至当前帧之后对当前帧应用去加重滤波器。在将该帧去加重(此去加重也充当对所添加噪声的低复杂性、陡峭IIR高通滤波)之后,音频解码器提供已解码音频信息。因此,根据图2的方法允许通过调整将要添加至当前帧的噪声的倾斜以改善背景噪声的质量来增强音频信息的声音质量。

[0043] 图3示出了根据本发明的音频解码器的第二实施方式。音频解码器同样适于基于已编码音频信息来提供已解码音频信息。音频解码器被配置为使用可基于AMR-WB、G.718及LD-USAC(EVS)的编码器来解码已编码音频信息。已编码音频信息同样包含可分别表示为系数 a_k 的线性预测系数(LPC)。根据第二实施方式的音频解码器包含:噪声水平估计器,其被配置为使用至少一个先前帧的线性预测系数来估计当前帧的噪声水平,以获得噪声水平信息;以及噪声插入器,其被配置为取决于由噪声水平估计器提供的噪声水平信息来将噪声添加至当前帧。噪声插入器被配置为在已编码音频信息的比特率小于每个样本0.5比特的条件下将噪声添加至当前帧。另外,噪声插入器可被配置为在当前帧为语音帧的条件下将噪声添加至当前帧。因此,同样可将噪声添加至当前帧以改善已解码音频信息的总体声音质量,该质量可因编码伪影而受损,尤其就语音信息的背景噪声而言。当考虑到至少一个先前音频帧的噪声水平来调整噪声的噪声水平时,可在不取决于位流中的边信息的情况下改善总体声音质量。因此,可减小将要用位流传递的数据的量。

[0044] 图4示出了根据本发明的用于执行音频解码的第二种方法,该方法可由根据图3的音频解码器执行。连同方法特征一起描述了图3中所描绘的音频解码器的技术细节。根据图4,音频解码器被配置为读取位流以便判定当前帧的帧类型。另外,音频解码器包含用于判定当前帧的帧类型的帧类型判定器,该帧类型判定器被配置为识别当前帧的帧类型为语音还是一般音频,使得可取决于当前帧的帧类型来执行噪声水平估计。一般而言,音频解码器适于:计算表示当前帧的未频谱整形的激发的第一信息,且计算关于当前帧的频谱缩放的第二信息以计算第一信息和第二信息之商来获得噪声水平信息。例如,若帧类型为ACELP(其是语音帧类型),则音频解码器解码当前帧的激发信号,且从该激发信号的时域表示来针对当前帧 f 计算其均方根 e_{rms} 。这意味着,音频解码器适于:在当前帧为语音类型的条件下,解码当前帧的激发信号,且从当前帧的时域表示(time domain representation)来计算其均方根 e_{rms} 来作为第一信息,以便获得噪声水平信息。在另一种情况下,若帧类型为MDCT或DTX(其是一般音频帧类型),则音频解码器解码当前帧的激发信号,且从该激发信号的时域表示等效物来针对当前帧 f 计算其均方根 e_{rms} 。这意味着,音频解码器适于:在当前帧为一般音频类型的条件下,解码当前帧的未整形的MDCT激发,且从当前帧的频谱域表示来计算其均方根 e_{rms} 来作为第一信息,以获得噪声水平信息。WO 2012/110476 A1中描述了具体如何完成上述操作。另外,图9示出了例示了如何从MDCT功率谱确定LPC滤波器等效物的图。虽然所描绘的尺度为巴克尺度,但也可从线性尺度获得LPC系数等效物。尤其当从线性尺度获得LPC系数等效物时,计算出的LPC系数等效物非常类似于根据例如以ACELP加以编

码的相同帧的时域表示所计算出的LPC系数。

[0045] 另外,如图4的方法图所例示,根据图3的音频解码器适于:在当前帧为语音类型的条件下,计算当前帧的LPC滤波器的传递函数的峰值水平 p 来作为第二信息,从而使用线性预测系数来获得噪声水平信息。这意味着,音频解码器根据公式 $p = \sum |a_k|$ 来计算当前帧的LPC分析滤波器的传递函数的峰值水平 p ,其中 a_k 为线性预测系数,其中 $k=0 \cdots 15$ 。若帧为一般音频信息,则从当前帧的频谱域表示获得LPC系数等效物,如图9所示以及W0 2012/110476 A1中及上文所描述的。如图4中所看出,在计算峰值水平 p 之后,通过将 e_{rms} 除以 p 来计算当前帧 f 的频谱最小值 m_f 。因此,音频解码器适于:计算表示当前帧的未频谱整形的激发的第一信息,该第一信息在此实施方式中为 e_{rms} ,且计算关于当前帧的频谱缩放的第二信息,该第二信息在此实施方式中为峰值水平 p ,以便计算第一信息和第二信息之商来获得噪声水平信息。然后在噪声水平估计器中将当前帧的频谱最小值加入队列,音频解码器适于:无论帧类型如何,在噪声水平估计器中将从当前音频帧获得的商加入队列,且噪声水平估计器包含用于从不同音频帧获得的两个或更多商(在此情况下为频谱最小值 m_f)的噪声水平寄存器。更具体而言,噪声水平寄存器可储存来自50个帧的商以便估计噪声水平。另外,噪声水平估计器适于基于不同音频帧的两个或更多商(因此,频谱最小值 m_f 的集合)的统计分析来估计噪声水平。在例示出必需的计算步骤的图7中详细描绘用于计算商 m_f 的步骤。在第二实施方式中,噪声水平估计器基于根据[3]已知的最小值统计来操作。若当前帧为语音帧,则根据基于最小值统计的当前帧的所估计噪声水平来缩放噪声,然后将噪声添加至当前帧。最后,将当前帧去加重(图4中未展示)。因此,此第二实施方式亦允许省略用于噪声填充的边信息,从而允许减小将要用位流传递的数据的量。因此,通过在解码阶段期间增强背景噪声而不增加数据速率,可改善音频信息的声音质量。请注意,因为无需时间/频率变换,且因为噪声水平估计器每个帧仅运行一次(而不是对多个子频带(sub-band)运行),所以所描述的噪声填充在能够改善有噪声的语音的低比特率编码的同时表现出极低的复杂性。

[0046] 图5示出了根据本发明的音频解码器的第三实施方式。

[0047] 音频解码器适于基于已编码音频信息来提供已解码音频信息。音频解码器被配置为使用基于LD-USAC的编码器来解码已编码音频信息。已编码音频信息包含可分别表示为系数 a_k 的线性预测系数(LPC)。音频解码器包含:倾斜调整器,其被配置为使用当前帧的线性预测系数来调整噪声的倾斜以获得倾斜信息;以及噪声水平估计器,其被配置为使用至少一个先前帧的线性预测系数来估计当前帧的噪声水平,以获得噪声水平信息。另外,音频解码器包含噪声插入器,其被配置为取决于通过倾斜计算器获得的倾斜信息且取决于通过噪声水平估计器提供的噪声水平信息来将噪声添加至当前帧。因此,取决于通过倾斜计算器获得的倾斜信息且取决于通过噪声水平估计器提供的噪声水平信息,可将噪声添加至当前帧以便改善解码后的音频信息的总体声音质量,该质量可因编码伪影而受损,尤其就语音信息的背景噪声而言。在此实施方式中,音频解码器所包含的随机噪声产生器(未展示)产生频谱白噪声,随后根据噪声水平信息来缩放该噪声并且使用 g 导出的倾斜对其加以整形,如先前所描述。

[0048] 图6示出了根据本发明的用于执行音频解码的第三种方法,该方法可由根据图5的音频解码器执行。读取位流,且被称为帧类型检测器的帧类型判定器判定当前帧为语音帧(ACELP)还是一般音频帧(TCX/MDCT)。无论帧类型如何,解码帧报头,且解码感知域

(perceptual domain)中的频谱整平后的(spectrally flattened)未整形的激发信号。在语音帧的情况下,此激发信号是时域激发,如先前所描述。若帧为一般音频帧,则解码MDCT域残余(频谱域)。分别使用时域表示及频谱域表示来估计噪声水平,如图7中所例示以及先前所描述的,从而使用也用来解码位流的LPC系数而不是使用任何边信息或额外的LPC系数。在当前帧为语音帧的条件下,将两种类型的帧的噪声信息加入队列,以调整将要添加至当前帧的噪声的倾斜和噪声水平。在将噪声添加至ACELP语音帧(应用ACELP噪声填充)之后,通过IIR将该ACELP语音帧去加重,且在表示已解码音频信息的时间信号中组合语音帧与一般音频帧。图6中通过小插图I、II及III描绘了去加重对所添加噪声的频谱的陡峭高通效应。

[0049] 换言之,根据图6,在LD-USAC (EVS) 解码器中实施上文所描述的ACELP噪声填充系统,该解码器是xHE-AAC[6]的低延迟变体,其可每个帧地在ACELP(语音)与MDCT(音乐/噪声)编码之间切换。将根据图6的插入过程概述如下:

[0050] 1. 读取位流,且判定当前帧为ACELP帧还是MDCT帧或DTX帧。无论帧类型如何,解码频谱整平后的激发信号(在感知域中)且将其用来更新噪声水平估计,如下文所详细描述那样。然后,直至为最后一个步骤的去加重,信号得以完全重新建构。

[0051] 2. 若帧经ACELP编码,则通过LPC过滤器系数的一阶LPC分析来计算用于噪声插入的倾斜(总体频谱形状)。该倾斜是从16个LPC系数 a_k 的增益 g 导出,增益 g 是由 $g = \sum [a_k \cdot a_{k+1}] / \sum [a_k \cdot a_k]$ 给出。

[0052] 3. 若帧经ACELP编码,则使用噪声整形水平及倾斜来执行对已解码帧的噪声添加:随机噪声产生器产生频谱白噪声信号,然后缩放该信号且使用 g 导出的倾斜对其加以整形。

[0053] 4. 紧接在最后的去加重填充步骤之前,将用于ACELP帧的已整形且已调平(levelled)的噪声信号添加至已解码信号。因为去加重是提升低频的一阶IIR,所以这允许对所添加噪声的低复杂性、陡峭IIR高通滤波,如同图6中一样,从而避免在低频处的可听见噪声伪影。

[0054] 步骤1中的噪声水平估计是通过以下操作来执行:计算当前帧的激发信号的均方根 e_{rms} (或在MDCT域激发的情况下为时域等效物,其意味着在帧为ACELP帧的情况下,将针对该帧来计算的 e_{rms}),以及随后将 e_{rms} 除以LPC分析滤波器的传递函数的峰值水平 p 。此操作得出帧 f 的频谱最小值的水平 m_f ,如同图7中一样。最后在基于例如最小值统计[3]来操作的噪声水平估计器中将 m_f 加入队列。请注意,因为不需要时间/频率变换,且因为该水平估计器每个帧仅运行一次(而不是对多个子频带运行),所以所描述的CELP噪声填充系统在能够改善有噪声的语音的低比特率编码的同时表现出极低的复杂性。

[0055] 虽然已就音频解码器为背景来描述了一些方面,但显然这些方面也表示对应的方法的描述,其中方块或设备对应于方法步骤或方法步骤的特征。类似地,就方法步骤为背景所描述的方面也表示对应的音频解码器的对应的方块或项目或特征的描述。该等方法步骤中的一些或全部可通过(或使用)例如为微处理器、可编程计算机或电子电路的硬件装置来执行。在一些实施方式中,最重要的方法步骤中的某一个或多个可通过这样的装置来执行。

[0056] 本发明的已编码音频信号可储存于数字储存介质上或可在传输介质上加以传输,传输介质为诸如无线传输介质或有线传输介质,诸如因特网。

[0057] 取决于特定的实行方案要求,本发明的实施方式可在硬件或软件中实行。可使用

储存有电子可读控制信号的数字储存介质来执行实行方案,数字储存介质例如软盘、DVD、蓝光盘、CD、ROM、PROM、EPROM、EEPROM或闪存,该等电子可读控制信号与可编程计算机系统合作(或能够与可编程计算机系统合作)以使得相应的方法得以执行。因此,数字储存介质可为计算机可读的。

[0058] 根据本发明的一些实施方式包含一种具有电子可读控制信号的数据载体,该等电子可读控制信号能够与可编程计算机系统合作以使得本文中所描述的方法之一得以执行。

[0059] 一般而言,本发明的实施方式可实行为一种具有程序代码的计算机程序产品,当该计算机程序产品在计算机上运行时,该程序代码可操作来执行该等方法中的一种。该程序代码可例如储存于机器可读载体上。

[0060] 其他实施方式包含用于执行本文中所描述的方法之一的计算机程序,其储存于机器可读载体上。

[0061] 换言之,本发明的方法的一实施方式因此是一种具有程序代码的计算机程序,当该计算机程序在计算机上运行时,该程序代码用于执行本文中所描述的方法之一。

[0062] 本发明方法的另一实施方式因此是一种数据载体(或数字储存介质或计算机可读媒体),其包含记录于其上的用于执行本文中所描述的方法之一的计算机程序。数据载体、数字储存介质或记录媒体通常为有形的和/或非暂时性的。

[0063] 本发明方法的另一实施方式因此是一种数据流或一种信号序列,其表示用于执行本文中所描述的方法之一的计算机程序。该数据流或该信号序列可例如被配置为经由数据通讯连接(例如经由因特网)加以传递。

[0064] 另一实施方式包含一种处理构件,例如计算机或可编程逻辑设备,其被配置为执行或适于执行本文中所描述的方法之一。

[0065] 另一实施方式包含一种计算机,其上安装有用于执行本文中所描述的方法之一的计算机程序。

[0066] 根据本发明的另一实施方式包含一种装置或一种系统,其被配置为将用于执行本文中所描述的方法之一的计算机程序(例如,电子地或光学地)传递至接收器。该接收器可例如为计算机、移动设备、内存设备或类似物。该装置或系统可例如包含用于将计算机程序传递至接收器的文件服务器。

[0067] 在一些实施方式中,可编程逻辑设备(例如场可编程门阵列)可用来执行本文中所描述的方法的功能中的一些或全部。在一些实施方式中,场可编程门阵列可与微处理器合作以便执行本文中所描述的方法之一。一般而言,较佳通过任何硬件装置来执行该等方法。

[0068] 可使用硬件装置,或使用计算机,或使用硬件装置与计算机之组合来实行本文中所描述之装置。

[0069] 可使用硬件装置,或使用计算机,或使用硬件装置与计算机之组合来实行本文中所描述的方法。

[0070] 上述实施方式仅例示出本发明的原理。应理解,本文中所描述的配置及细节的修改及变化对本领域技术人员而言将显而易见。因此,将仅受申请专利权利要求书的范围的限制,而不受本文中对实施方式的描述及说明所呈现的特定细节限制。

[0071] 非专利文献引用清单

[0072] [1]B.Bessette et al.,“The Adaptive Multi-rate Wideband Speech Codec

(AMR-WB),”IEEE Trans.On Speech and Audio Processing,Vol.10,No.8,Nov.2002。

[0073] [2]R.C.Hendriks,R.Heusdens and J.Jensen,“MMSE based noise PSD tracking with low complexity,”in IEEE Int.Conf.Acoust.,Speech,Signal Processing,pp.4266-4269,March 2010。

[0074] [3]R.Martin,“Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics,”IEEE Trans.On Speech and Audio Processing, Vol.9,No.5,Jul.2001。

[0075] [4]M.Jelinek and R.Salami,“Wideband Speech Coding Advances in VMR-WB Standard,”IEEE Trans.On Audio,Speech,and Language Processing,Vol.15,No.4,May 2007。

[0076] [5]J. **Mäkinen** et al.,“AMR-WB+:A New Audio Coding Standard for 3rd Generation Mobile Audio Services,”in Proc.ICASSP 2005,Philadelphia,USA, Mar.2005。

[0077] [6]M.Neuendorf et al.,“MPEG Unified Speech and Audio Coding-The ISO/MPEG Standard for High-Efficiency Audio Coding of All Content Types,”in Proc.132nd AES Convention,Budapest,Hungary,Apr.2012.Also appears in the Journal of the AES,2013。

[0078] [7]T.Vaillancourt et al.,“ITU-T EV-VBR:A Robust 8-32kbit/s Scalable Coder for Error Prone Telecommunications Channels,”in Proc.EUSIPCO 2008, Lausanne,Switzerland,Aug.2008。

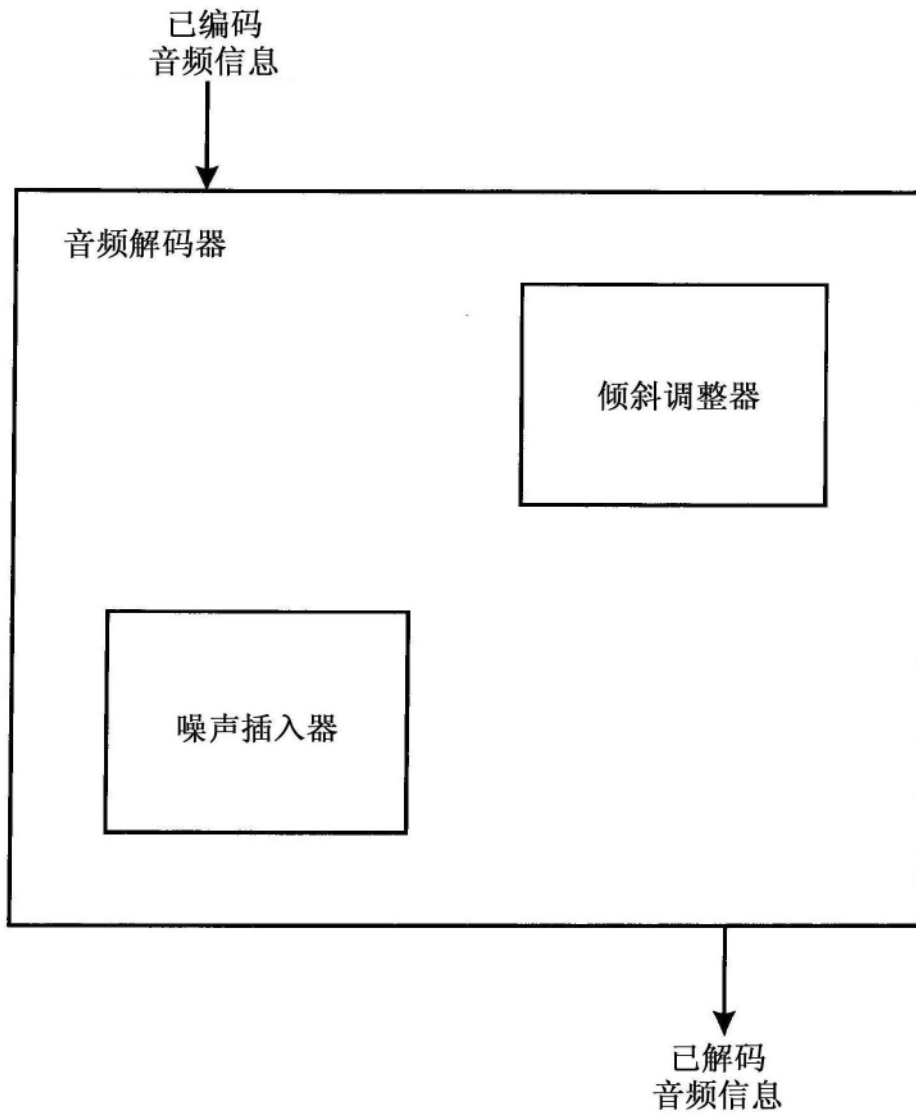


图1

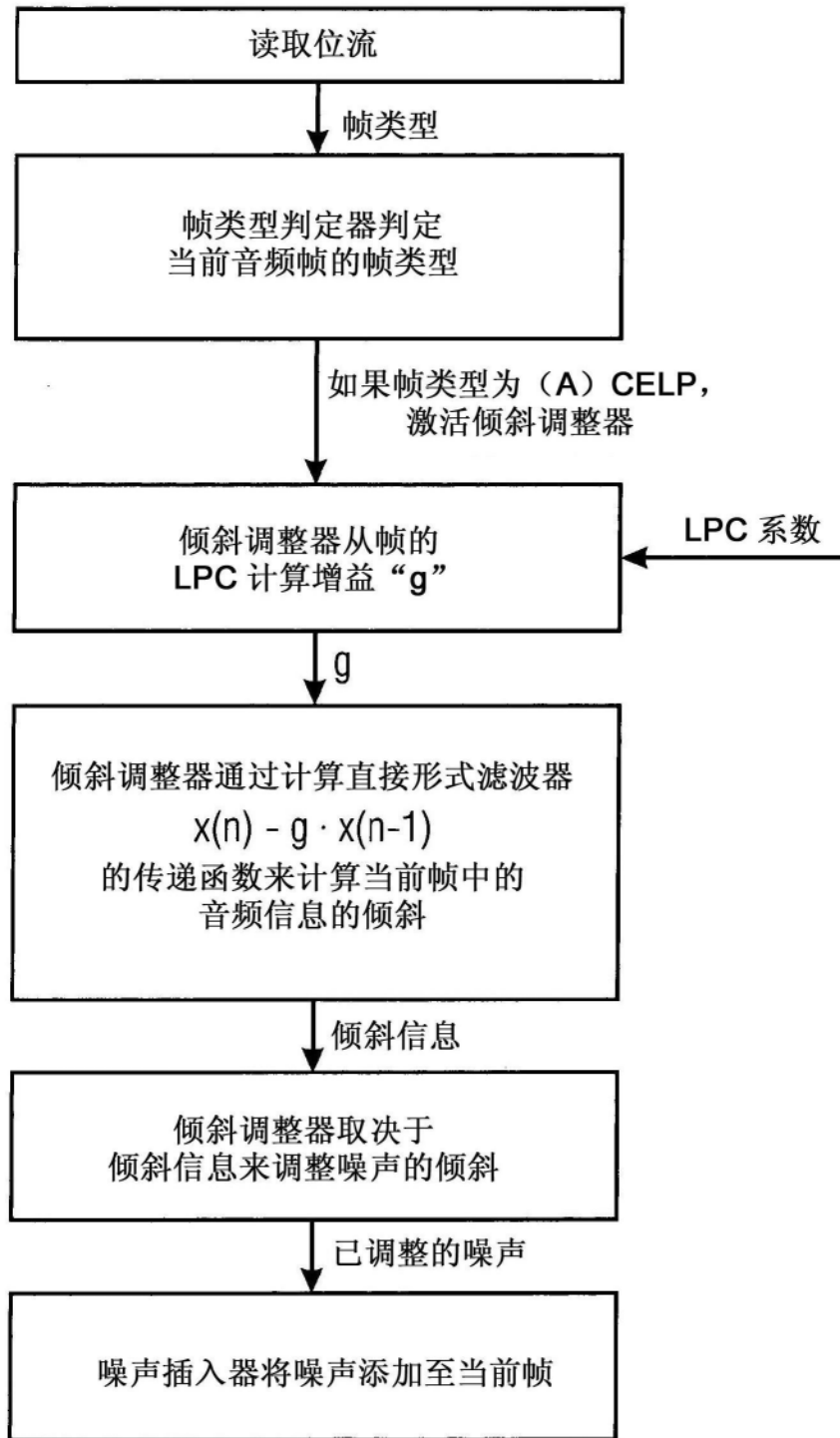


图2

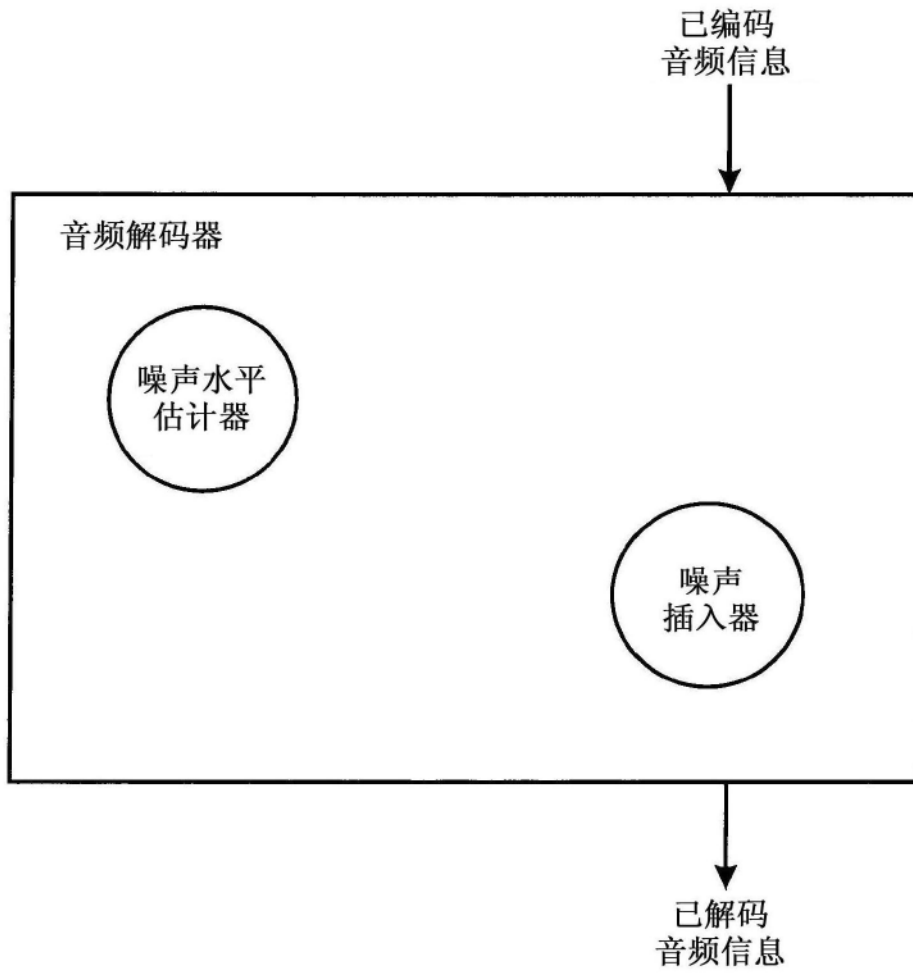


图3

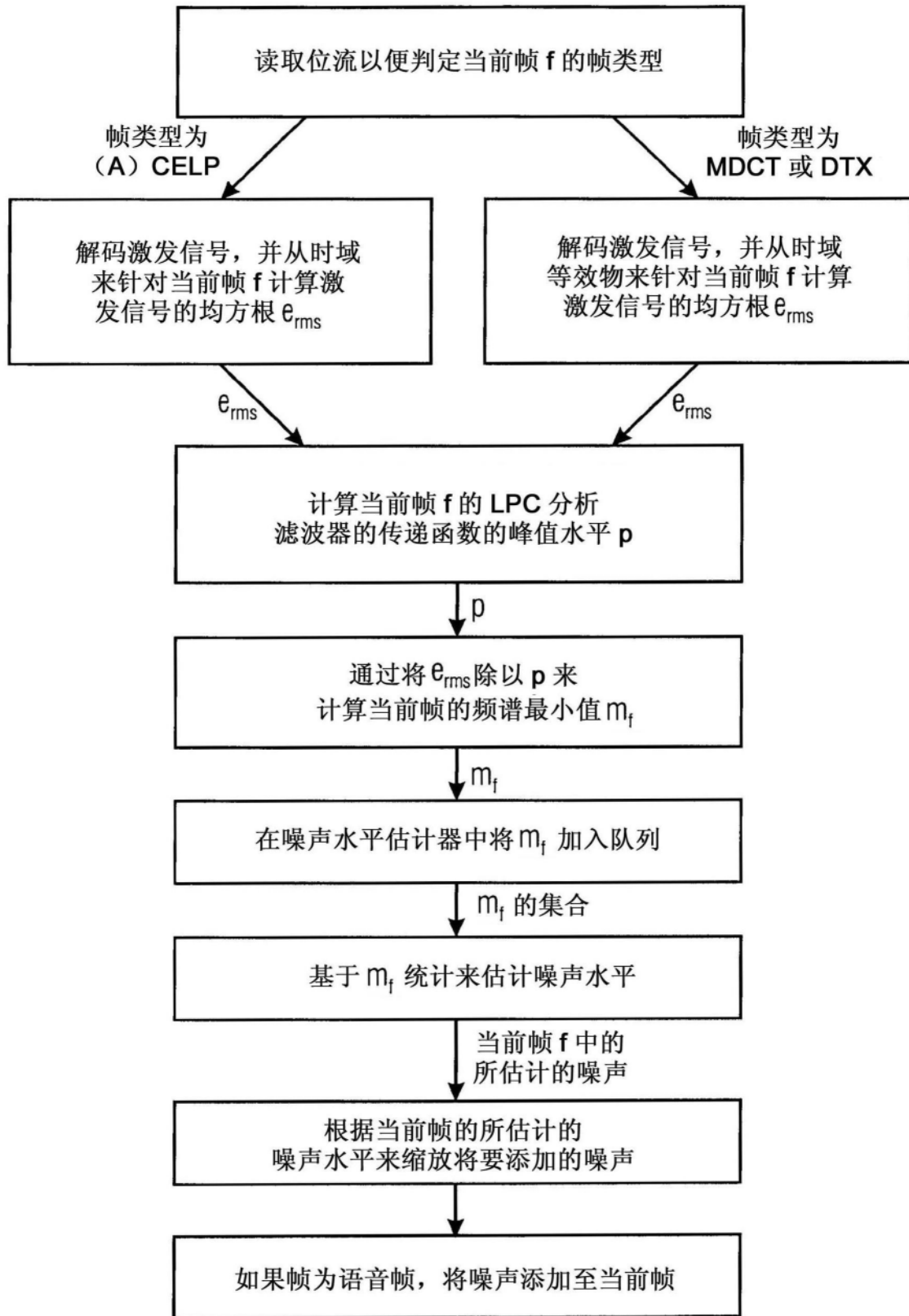


图4

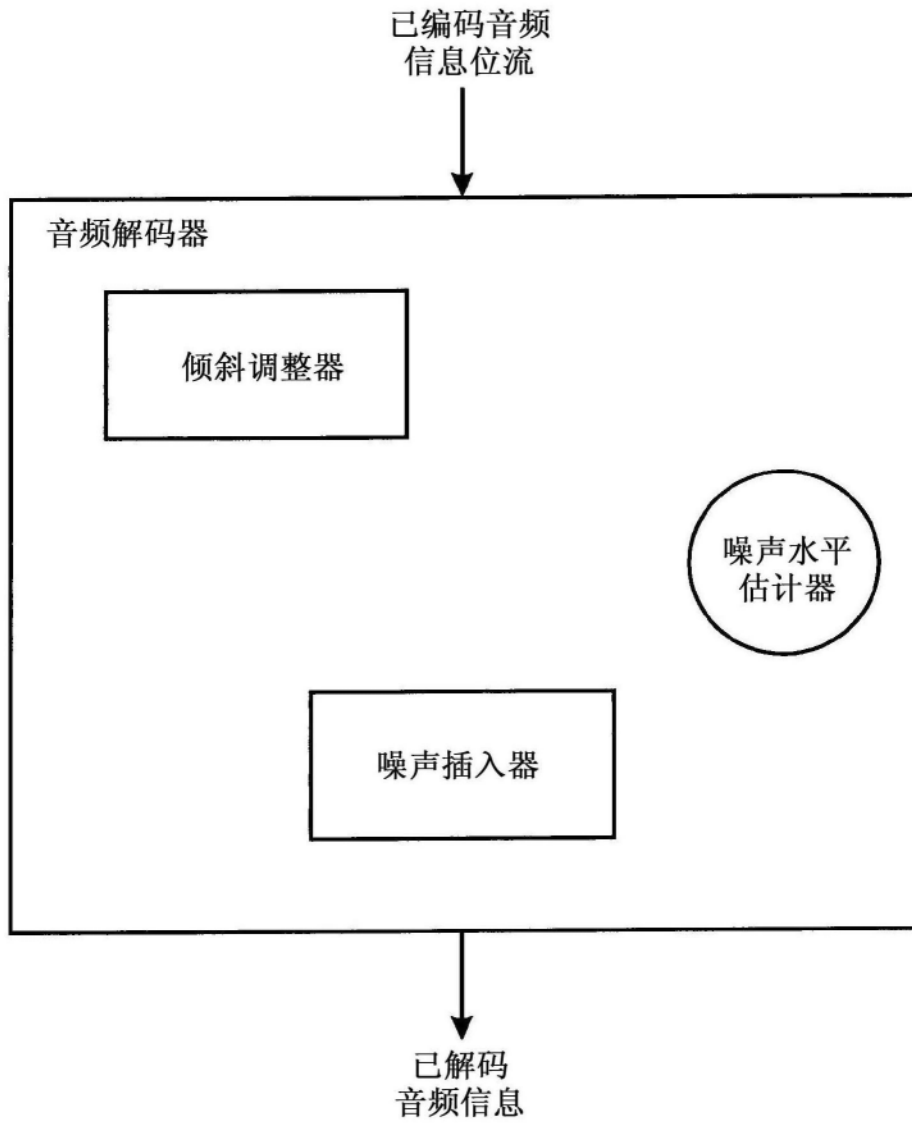


图5

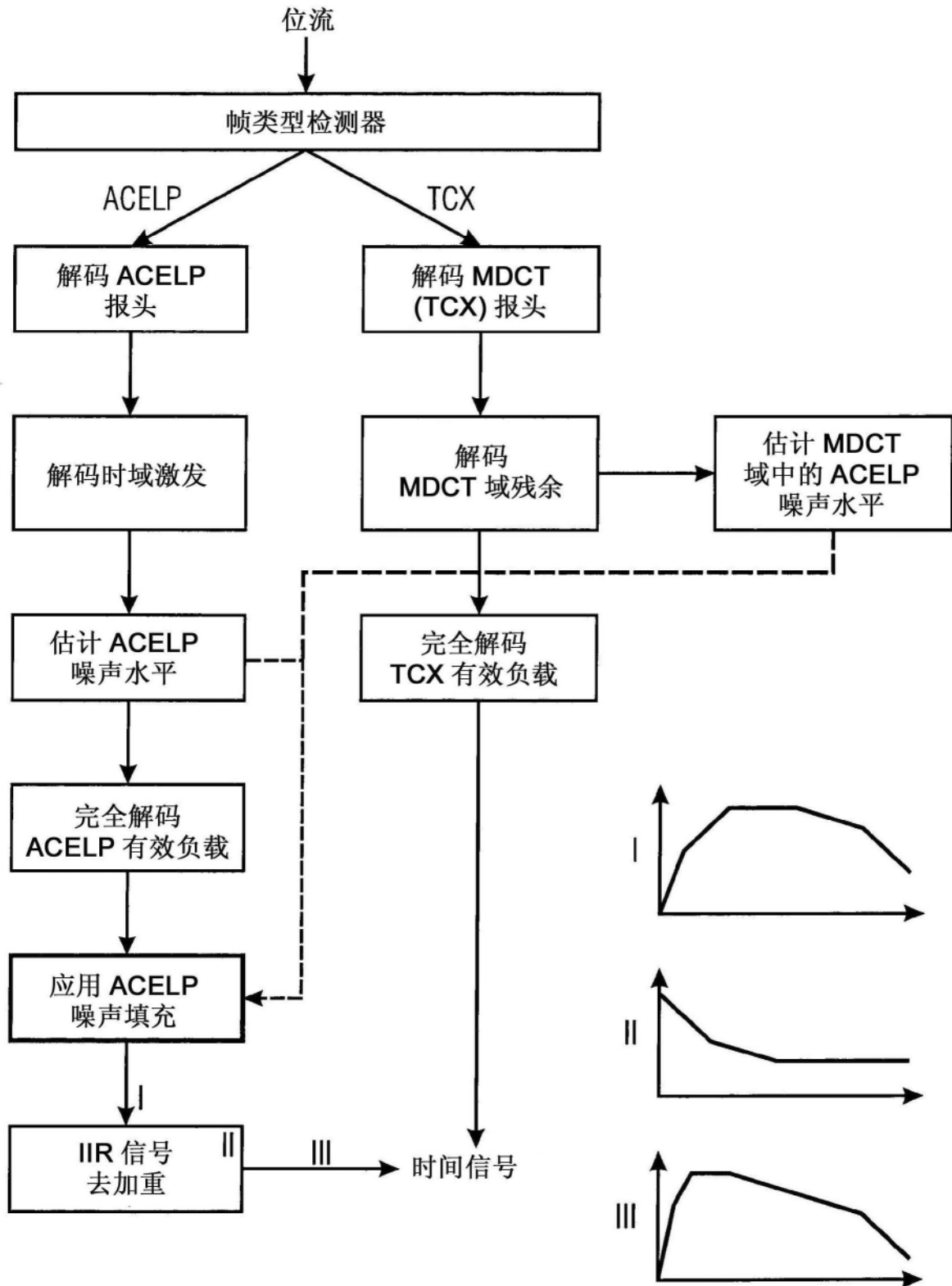
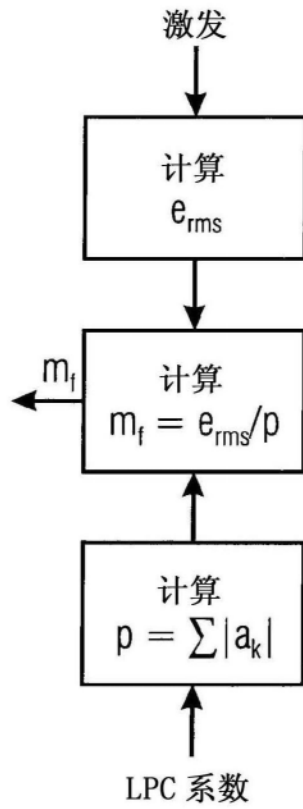
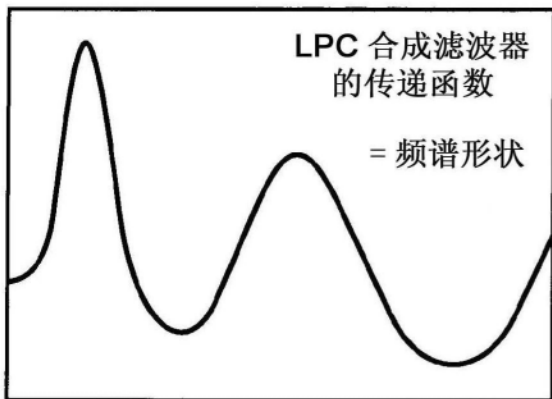


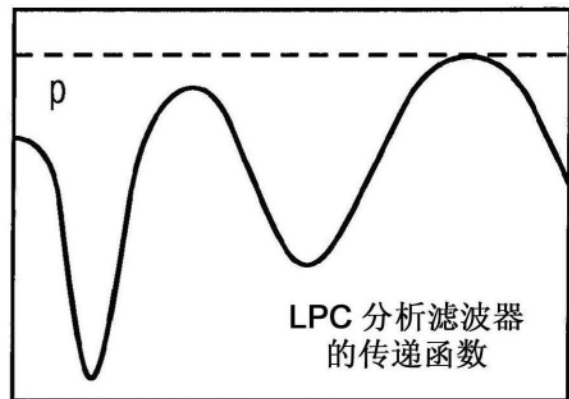
图6



A

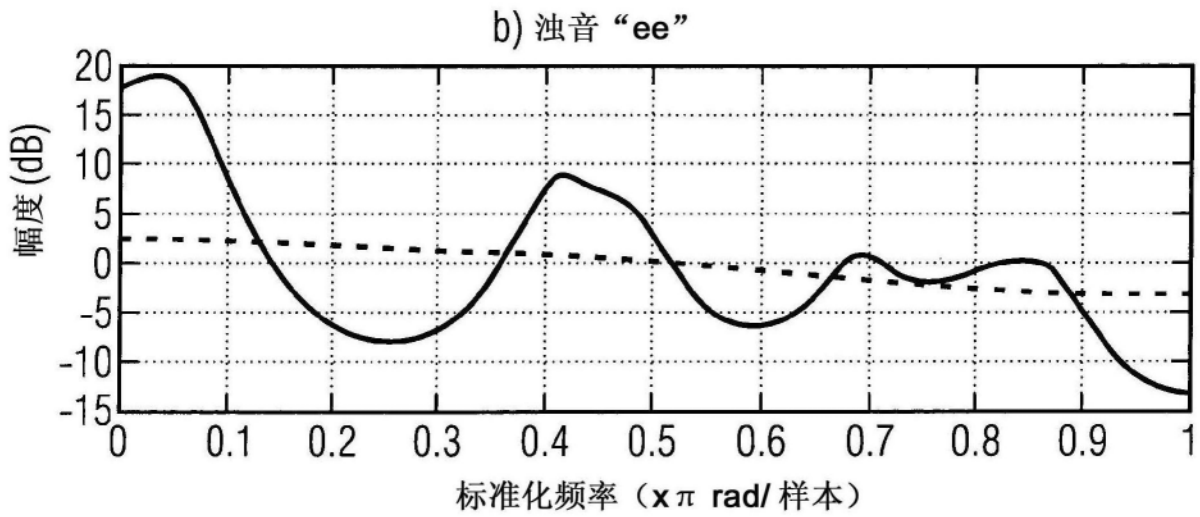
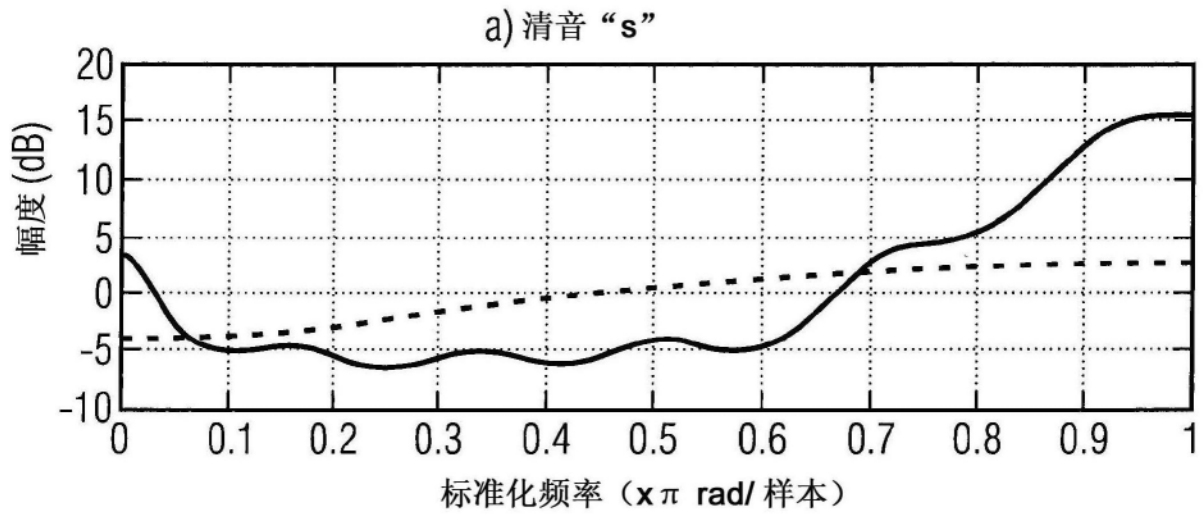


B



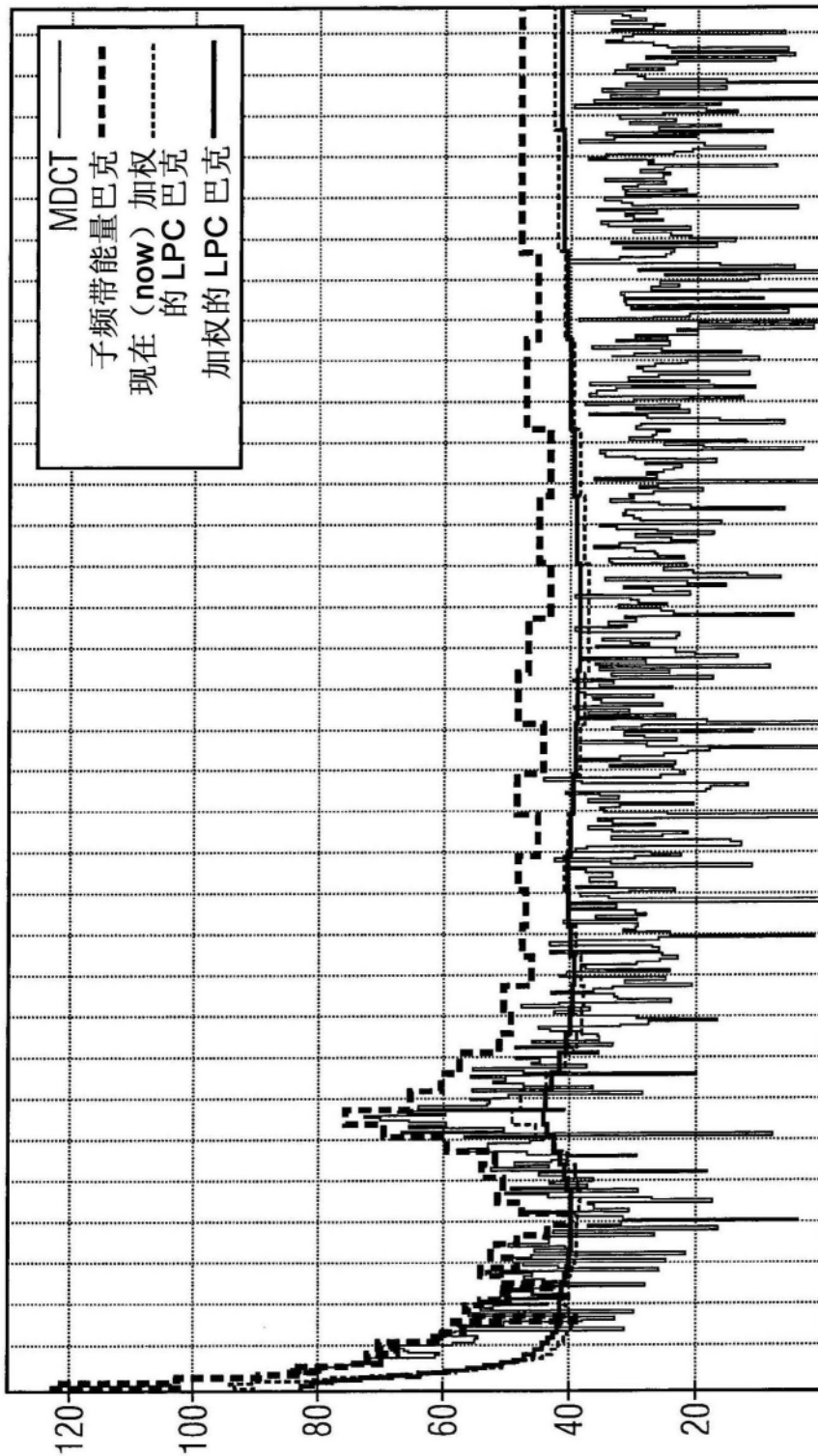
C

图7



从 LPC 系数导出的 (- - -) 频谱倾斜的,
(——)LPC 滤波器的传递函数

图8



从 MDCT 功率谱 (MDCT 值的平方) 确定 LPC 滤波器

图9