



(21)申请号 201480047702.3

(22)申请日 2014.09.09

(65)同一申请的已公布的文献号
申请公布号 CN 105637583 A

(43)申请公布日 2016.06.01

(30)优先权数据
61/875,690 2013.09.10 US
14/478,839 2014.09.05 US

(85)PCT国际申请进入国家阶段日
2016.02.29

(86)PCT国际申请的申请数据
PCT/CN2014/086135 2014.09.09

(87)PCT国际申请的公布数据
W02015/035896 EN 2015.03.19

(73)专利权人 华为技术有限公司

地址 518129 广东省深圳市龙岗区坂田华为总部办公楼

(72)发明人 高扬

(51)Int.Cl.
G10L 19/032(2006.01)

(56)对比文件
CN 103026408 A,2013.04.03,
CN 1185626 C,2005.01.19,
CN 1496559 A,2004.05.12,
WO 2013/035257 A1,2013.03.14,
CN 101273404 A,2008.09.24,
CN 102044250 A,2011.05.04,

审查员 游晓梅

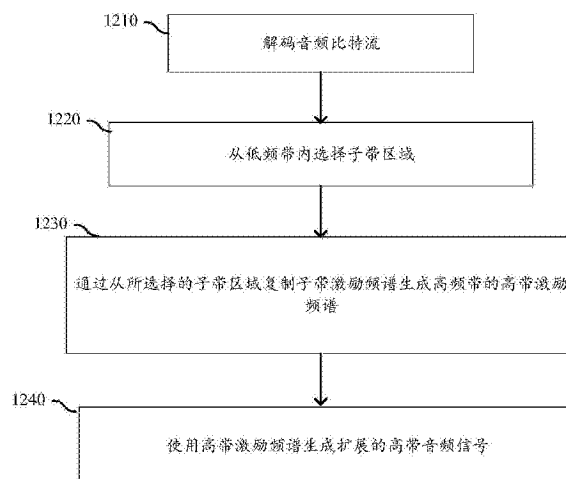
权利要求书2页 说明书13页 附图13页

(54)发明名称

自适应带宽扩展方法及其装置

(57)摘要

在本发明的一项实施例中,一种解码已编码音频比特流和生成频带扩展的方法包括解码所述音频比特流以产生已解码低带音频信号和生成对应于低频带的低带激励频谱。使用参数从所述低频带内选择子带区域,所述参数指示所述已解码低带音频信号的频谱包络的能量信息。通过从所述选择的子带区域复制子频带激励频谱到对应于高频带的高子带区域生成所述高频带的高带激励频谱。使用所述生成的高带激励频谱,通过采用高带频谱包络生成扩展的高带音频信号。将所述扩展的高带音频信号添加到所述已解码低带音频信号以生成具有扩展的频率带宽的音频输出信号。



1. 一种在解码器处对已编码音频比特流进行解码以及生成频带扩展的方法,其特征在于,所述方法包括:

对所述音频比特流进行解码以产生已解码低带音频信号以及生成对应于低频带的低带激励频谱;

使用参数从所述低频带内选择子带区域,所述参数指示所述已解码低带音频信号的频谱包络的能量信息;其中,所述选择的子带区域为对应于最高频谱包络能量的子带区域;

通过从所述选择的子带区域内复制子带激励频谱到对应于高频带的高子带区域以生成所述高频带的高带激励频谱;

使用所述生成的高带激励频谱和高带频谱包络生成扩展的高带音频信号;以及

将所述扩展的高带音频信号添加到所述已解码低带音频信号以生成具有扩展的频率带宽的音频输出信号。

2. 根据权利要求1所述的方法,其特征在于,指示所述频谱包络的能量信息的所述参数包括反映所述频谱包络的最高能量或频谱共振峰峰值的参数。

3. 根据权利要求1或2所述的方法,其特征在于,所述解码方法采用带宽扩展技术以生成所述高频带。

4. 根据权利要求1或2所述的方法,其特征在于,采用所述高带频谱包络包括采用表示所述高带频谱包络的预测的高带滤波器。

5. 根据权利要求1或2所述的方法,其特征在于,还包括:

通过将所述频域音频频谱反变换为时域以生成所述音频输出信号。

6. 根据权利要求1或2所述的方法,其特征在于,从所述选择的子带区域内复制所述子带激励频谱到对应于所述高频带的所述高子带区域包括从滤波器组分析复制输出的低频带系数到高子带区域。

7. 根据权利要求1或2所述的方法,其特征在于,所述音频比特流包括浊音语音或谐波音乐。

8. 一种对已编码音频比特流进行解码和生成频谱带宽的解码器,其特征在于,所述解码器包括:

低带解码单元,用于对所述音频比特流进行解码以产生已解码低带音频信号和生成对应低频带的低频带激励频谱;以及

带宽扩展单元,耦合到所述低带解码单元并且包括子带选择单元和复制单元,其中所述子带选择单元用于使用参数从所述低频带内选择子带区域,所述参数指示所述已解码低带音频信号的频谱包络的能量信息,所述复制单元用于通过从所述选择的子带区域内复制子带激励频谱到对应于高频带的高子带区域以生成所述高频带的高带激励频谱,其中,所述选择的子带区域为对应于最高频谱包络能量的子带区域。

9. 根据权利要求8所述的解码器,其特征在于,指示所述频谱包络的能量信息的所述参数包括反映频谱包络能量或频谱共振峰峰值的参数。

10. 根据权利要求8或9所述的解码器,其特征在于,还包括:

耦合到所述复制单元的高带信号生成器,所述高带信号生成器用于使用预测的高带频谱包络生成高带时域信号;以及

耦合到所述高带信号生成器和所述低带解码单元的输出生成器,其中所述输出生成器

用于通过组合由解码所述音频比特流获得的低带时域信号与所述高带时域信号以生成音频输出信号。

11. 根据权利要求10所述的解码器,其特征在于,所述高带信号生成器用于使用表示所述预测的高带频谱包络的预测的高带滤波器。

12. 根据权利要求8或9所述的解码器,其特征在于,还包括:

耦合到所述复制单元的高带频谱生成器,所述高带频谱生成器用于使用估计的高带频谱包络和所述高带激励频谱生成所述高频带的高带频谱;以及

耦合到所述高带频谱生成器和所述低带解码单元的输出频谱生成器,其中所述输出频谱生成器用于通过组合由解码所述音频比特流获得的低带频谱与所述高带频谱生成频域音频频谱。

13. 根据权利要求12所述的解码器,其特征在于,还包括:

反变换信号生成器,用于通过将所述频域音频频谱反变换到时域生成时域音频信号。

14. 一种语音处理解码器,其特征在于,包括:

处理器,;以及

存储由所述处理器执行的程序的计算机可读存储介质,所述处理器被配置为执行所述计算机可读存储介质中存储的程序,以用于:

对所述音频比特率进行解码以产生已解码低带音频信号和生成对应于低频带的低带激励频谱;

使用参数从所述低频带内选择子带区域,所述参数指示所述已解码低带音频信号的频谱包络的能量信息,其中,所述选择的子带区域为对应于最高频谱包络能量的子带区域;

通过从所述选择的子带区域内复制子带激励频谱到对应于高频带的高子带区域生成所述高频带的高带激励频谱;

使用所述生成的高带激励频谱和高带频谱包络生成扩展的高带音频信号,以及

将所述扩展的高带音频信号添加到所述已解码低带音频信号以生成具有扩展的频率带宽的音频输出信号。

15. 根据权利要求14所述的解码器,其特征在于,指示所述频谱包络的能量信息的所述参数包括反映频谱包络能量或频谱共振峰峰值的参数。

自适应带宽扩展方法及其装置

[0001] 本发明要求2014年9月5日递交的发明名称为“自适应带宽扩展方法及其装置 (Adaptive Bandwidth Extension and Apparatus for the Same)”的第14/478,839号美国专利申请案的在先申请优先权,该在先申请是2013年9月10日递交的发明名称为“基于带宽扩展的频谱能量级的位移频带的自适应选择 (Adaptive Selection of Shifting Band Based on Spectral Energy Level for Bandwidth Extension)”的第61/875,690号美国临时专利申请案的连续申请案,这两个在先申请的内容以引入的方式并入本文本中,如全文再现一般。

技术领域

[0002] 本发明大体上涉及语音处理领域,尤其涉及自适应带宽扩展方法及其装置。

背景技术

[0003] 在现代音频/语音数字信号通信系统中,数字信号在编码器处被压缩,已压缩的信息(比特流)可以打包并且通过通信信道逐帧向解码器发送。编码器和解码器在一起的系统被称为编解码器。语音/音频压缩可以用来减少表示语音/音频信号的比特数,从而降低传输所需的比特率。语音/音频压缩技术大体上可以分为时域编码和频域编码。时域编码通常用于编码低比特率的语音信号或音频信号。频域编码通常用于编码高比特率的音频信号或语音信号。带宽扩展(BWE)可以是时域编码或频域编码的一部分,用于以非常低的比特率或以零比特率生成高带信号。

[0004] 然而,语音编码器是有损编码器,即,解码得到信号不同于原始信号。因此,语音编码的目标之一是为了使失真(或可感知损失)在给定比特率下最小化,或者使比特率最小化以达到给定的失真。

[0005] 语音编码与其它形式的音频编码的不同之处在于语音是一种比大多数其它音频信号简单地多的信号,而且关于语音特性的统计信息更多。因此,与音频编码相关的一些听觉信息在语音编码上下文中可以是不必要的。在语音编码中,最重要的标准是在传输的数据量受限的情况下保持语音的清晰度和“愉悦度”。

[0006] 语音的清晰度,除了包括实际文字内容,还包括说话者身份、情绪、语调、音色,所有这些对于最佳清晰度都很重要。受损语音的愉悦度是一个较抽象的概念,它不同于清晰度的一个特性,因为退化语音有可能是完全清晰的,但是主观上另听众厌烦。

[0007] 语音波形的冗余与不同类型的语音信号有关,例如浊音和清音语音信号。浊音,例如‘a’、‘b’,基本上是由于声带的振动而产生的,而且是振荡的。因此,在一个较短时间内,通过正弦等周期性信号的叠加可以很好地模拟它们。换言之,浊音语音信号基本上是周期性的。然而,这种周期性在语音片段的持续时间内可能是变化的,而且周期性波的形状通常随着片段而逐渐地变化。低比特率语音编码可以大大受益于研究这种周期性。浊音语音周期还被称为基音(pitch),基音预测通常被称为长期预测(LTP)。相比之下,清音,例如‘s’、‘sh’,更像噪声。这是因为清音语音信号更像一种随机噪声,并且具有更小的可预测性。

[0008] 传统上,所有参数语音编码方法利用语音信号内在的冗余来减少发送的信息量以及在短的间隔内估计信号的语音样本的参数。这种冗余主要是因为语音波形以准周期速率重复,以及语音信号的频谱包络变化缓慢。

[0009] 可以参考若干不同类型的语音信号,例如浊音和清音,考虑语音波形的冗余度。尽管浊音语音信号基本上是周期性的,但是这种周期性在语音片段的持续时间内可能是变化的,而且周期性波的形状通常随着片段而逐渐地变化。低比特率语音编码可以大大受益于研究这种周期性。浊音语音周期还被称为基音,基音预测通常被称为长期预测(LTP)。至于清音语音,信号更像一种随机噪声,并且具有更小的可预测性。

[0010] 在任一情况下,参数编码可以用于通过将语音信号的激励分量与频谱包络分量分离来减少语音片段的冗余。缓慢变化的频谱包络可以通过线性预测编码(LPC),也称为短期预测(STP)表示。低比特率语音编码还可以大大受益于研究此种短期预测。编码的优势来自于参数的缓慢变化。然而,这些参数与在几毫秒内保持的值明显不同是很少见的。相应地,在8kHz、12.8kHz或16kHz的采样速率下,语音编码算法采用的标称帧持续时间的范围在十到三十毫秒内。20毫秒的帧持续时间是最常见的选择。

[0011] 基于滤波器组技术的音频编码被广泛使用,例如在频域编码中。在信号处理中,滤波器组是一组将输入信号分离为多个分量的带通滤波器,每个带通滤波器携带原始信号的单个子频带。由滤波器组执行的解压缩过程被称为分析,而滤波器组分析的输出被称为子带信号,其中子带信号具有和滤波器组中的滤波器个数一样多的子带。重构过程被称为滤波器组合成。在数字信号处理中,术语“滤波器组”通常还应用于接收器组。区别在于接收器还将子带下转换成可以以更低的速率重新采样的低中心频率。有时可以通过对带通子带进行下采样获得相同的结果。滤波器组分析的输出可以采用复合系数形式。每个复合系数包含分别表示滤波器组中每个子带的余弦项和正弦项的实元素和虚元素。

[0012] 在最近的著名标准,例如G.723.1、G.729、G.718、增强型全速率(EFR)、可选择模式声码器(SMV)、自适应多速率(AMR)、可变速率多模式宽带(VMR-WB),或自适应多速率宽带(AMR-WB)中,已经采用了码激励线性预测技术(“CELP”)。CELP通常理解为编码激励、长期预测和短期预测的技术组合。CELP主要利用人类声音特性或人类嗓音发声模型对语音信号进行编码。CELP语音编码在语音压缩领域是一种非常普遍的算法原理,尽管不同编解码器中的CELP细节可能有很大不同。由于它的普遍性,CELP算法已经应用于ITU-T、MPEG、3GPP和3GPP2等各种标准中。CELP的变体包括代数CELP、广义CELP、低时延CELP和矢量和激励线性预测,以及其它。CELP是一类算法的通用术语,而不是针对特定的编解码器。

[0013] CELP算法基于四个主要观点。第一,使用通过线性预测(LP)的语音生成的源滤波器模型。用于语音生成的源滤波器将语音模拟为声源,例如声带,和线性声滤波器,即声道(和辐射特性)的组合。在语音生成的源滤波器模型的实施方式中,声源或激励信号通常被模拟为浊音语音的周期性脉冲序列,或清音语音的白噪声。第二,将自适应和固定码本用作LP模型的输入(激励)。第三,在“感知加权域”的闭环中执行搜索。第四,使用了矢量量化(VQ)。

发明内容

[0014] 本发明实施例描述一种在解码器处对已编码音频比特流进行解码以及生成频带

扩展的方法。所述方法包括对所述音频比特流进行解码以产生已解码低带音频信号以及生成对应于低频带的低带激励频谱。使用参数从所述低频带内选择子带区域,所述参数指示所述已解码低带音频信号的频谱包络的能量信息。通过从所述选择的子带区域复制子带激励频谱到对应于高频带的高子带区域生成所述高频带的高带激励频谱。使用所述生成的高带激励频谱通过采用高带频谱包络生成扩展的高带音频信号。将所述扩展的高带音频信号添加到所述已解码低带音频信号以生成具有扩展的频率带宽的音频输出信号。

[0015] 根据本发明的一个替代性实施例,一种用于对已编码音频比特流进行解码和生成频率带宽的解码器包括低带解码单元,用于解码所述音频比特流以产生已解码低带音频信号以及生成对应于低频带的低带激励频谱。所述解码器还包括耦合到所述低带解码单元的带宽扩展单元。所述带宽扩展单元包括子带选择单元和复制单元。所述子带选择单元用于使用指示所述已解码低带音频信号的频谱包络的能量信息的参数从所述低频带内选择子带区域。所述复制单元用于通过从所选择的子带区域复制子带激励频谱到对应于高频带的高子带区域生成所述高频带的高带激励频谱。

[0016] 根据本发明的一个替代性实施例,一种用于语音处理的解码器包括处理器和存储由所述处理器执行的程序的计算机可读存储介质。所述程序包括执行以下操作的指令:对所述音频比特流进行解码以产生已解码低带音频信号以及生成对应于低频带的低带激励频谱。所述程序包括执行以下操作的指令:使用参数从所述低频带内选择子带区域,所述参数指示所述已解码低带音频信号的频谱包络的能量信息;以及通过从所述选择的子带区域复制子带激励频谱到对应于高频带的高子带区域生成所述高频带的高带激励频谱。所述程序还包括执行以下操作的指令:使用所述生成的高带激励频谱通过采用高带频谱包络生成扩展的高带音频信号,以及将所述扩展的高带音频信号添加到所述已解码低带音频信号以生成具有扩展的频率带宽的音频输出信号。

[0017] 本发明的一替代性实施例描述一种在解码器处对已编码音频比特流进行解码以及生成频带扩展的方法。所述方法包括对所述音频比特流进行解码以产生已解码低带音频信号和生成对应于低频带的低带频谱,以及使用参数从所述低频带内选择子带区域,所述参数指示所述已解码低带音频信号的频谱包络的能量信息。所述方法还包括通过从所述选择的子带区域复制子带频谱到高频带区域生成高带频谱,以及使用所述生成的高带频谱以通过采用高带频谱包络能量生成扩展的高带音频信号。所述方法还包括将所述扩展的高带音频信号添加到所述已解码低带音频信号以生成具有扩展的频率带宽的音频输出信号。

附图说明

[0018] 为了更完整地理解本发明及其优点,现在参考下文结合附图进行的描述,其中:

[0019] 图1示出了使用传统CELP编码器对原始语音进行编码期间执行的操作;

[0020] 图2示出了下文描述的本发明实施例中传统CELP解码器对原始语音进行解码期间执行的操作;

[0021] 图3示出了在传统CELP编码器中对原始语音进行编码期间执行的操作;

[0022] 图4示出了实施如将下文所述的本发明实施例中对应于图5中的编码器的基本CELP解码器;

[0023] 图5A和5B示出了利用带宽扩展(BWE)的编码/解码的示例,其中图5A示出了具有

BWE边信息的编码器处的操作,而图5B示出了具有BWE的解码器处的操作;

[0024] 图6A和6B示出了在没有传输边信息的情况下利用BWE的编码/解码的另一示例,其中图6A示出了编码器处的操作,而图6B示出了解码器处的操作;

[0025] 图7示出了使用CELP类型的编解码器时的浊音语音或谐波音乐的理想化激励频谱的示例;

[0026] 图8示出了使用CELP类型的编解码器时的浊音语音或谐波音乐的已解码激励频谱的传统带宽扩展的示例;

[0027] 图9示出了本发明实施例使用CELP类型的编解码器时的应用于浊音语音或谐波音乐的已解码激励频谱的带宽扩展的示例;

[0028] 图10示出了本发明实施例中用于为BWE实施子带位移或复制的的解码器处的操作;

[0029] 图11示出了用于为BWE实施子带位移或复制的解码器的替代性实施例;

[0030] 图12示出了根据本发明实施例的解码器执行的操作;

[0031] 图13A和13B示出了根据本发明实施例的用于实施带宽扩展的解码器;

[0032] 图14示出了根据本发明实施例的通信系统;以及

[0033] 图15示出了可以用于实施本文公开的设备和方法的处理系统的框图。

具体实施方式

[0034] 在现代音频/语音数字信号通信系统中,数字信号在编码器处压缩,已压缩信息或比特流可以打包并且通过通信信道逐帧向解码器发送。解码器接收和解码已压缩信息以获取音频/语音数字信号。

[0035] 本发明大体上涉及语音/音频信号编码和语音/音频信号带宽扩展。尤其是,本发明实施例可用于改进带宽扩展领域中的ITU-T AMR-WB语音编码器的标准。

[0036] 一些频率比其它频率重要。这些重要的频率以高分辨率进行编码。这些频率之间的细微差别是很重要的,因此需要能保持这些差别的编码方案。另一方面,不那么重要的频率不必精确。可以使用更粗略的编码方案,即使一些更精细的细节将在编码时会丢失。典型的更粗略的编码方案是基于带宽扩展(BWE)的概念。这种技术概念还被称为高带扩展(HBE)、子带复制(SBR)或频带复制(SBR)。尽管名称可能不同,但是它们都具有相同的意义,即,利用非常低的比特率(甚至是零比特率)或明显低于正常编码/解码方法的比特率对一些子频带(通常是高带)进行编码/解码。

[0037] 在SBR技术中,可以从低频带复制高频带中的频谱精细结构,而且可以添加一些随机噪声。随后,通过使用从编码器向解码器传输的边信息形成高频带中的频谱包络。从低带到高带的频带位移或复制通常是BWE技术的第一步。

[0038] 本发明实施例将描述基于频谱包络的能量等级自适应选择位移频带来改进BWE的技术。

[0039] 图1示出了在使用传统CELP编码器对原始语音进行编码期间执行的操作。

[0040] 图1示出了传统初始CELP编码器,其中通常通过使用合成分析方法使合成语音102和原始语音101之间的加权误差109最小化,这意味着通过在闭环中感知优化已解码(合成)信号以执行编码(分析)。

[0041] 所有语音编码器利用的基本原理是语音信号是高度相关的波形这一事实。作为说明,可以使用如下公式(11)所示的自回归(AR)模型表示语音。

$$[0042] \quad X_n = \sum_{i=1}^L a_i X_{n-i} + e_n \quad (11)$$

[0043] 在公式(11)中,每个样本被表示为前L个样本加白噪声的线性组合。加权系数 a_1 、 a_2 …… a_L 被称为线性预测系数(LPC)。对于每个帧,选择加权系数 a_1 、 a_2 …… a_L ,使得使用上述模型生成的频谱 $\{X_1, X_2, \dots, X_N\}$ 最匹配输入语音帧的频谱。

[0044] 可选地,语音信号还可以通过谐波模型和噪声模型的组合表示。模型的谐波部分实际上是信号的周期性分量的傅里叶级数表示。一般而言,对于浊音信号,语音的谐波加噪声模型由谐波和噪声的混合物组成。浊音语音中的谐波和噪声的比例取决于多个因素,包括说话者特征(例如,说话者的声音在哪个程度上是正常的还是像呼吸声);语音片段特征(例如,语音片段在哪个程度上是周期性的)以及频率。浊音语音的较高频率具有较高比例的噪声型分量。

[0045] 线性预测模型和谐波噪声模型是用于对语音信号进行模拟和编码的两个主要方法。线性预测模型尤其擅长对语音的频谱包络进行模拟,而谐波噪声模型擅长对语音的精细结构进行模拟。可以组合这两个方法以利用它们的相对强度。

[0046] 如先前所指示,在进行CELP编码之前,例如以每秒8000个样本的速率,对到达手机麦克风的输入信号进行滤波和采样。随后,例如利用每个样本13个比特对每个样本进行量化。将采样的语音分段成20ms的片段或帧(例如,在160个样本的情况下)。

[0047] 分析语音信号,并且提取它的LP模型、激励信号和基音。LP模型表示语音的频谱包络。它被转换到一组线谱频率(LSF)系数,其是线性预测参数的替代性表示,因为LSF系数具有良好的量化特性。可以对LSF系数进行标量量化,或者更高效地,可以使用先前训练的LSF矢量码本对它们进行矢量量化。

[0048] 码激励包括含有码矢量的码本,这些码矢量具有全部独立选择的分量,使得每个码矢量可以具有近似‘白’频谱。对于输入语音的每个子帧而言,通过短期线性预测滤波器103和长期预测滤波器105对每个码矢量进行滤波,并且将输出与语音样本进行比较。在每个子帧处,选择输出最佳匹配输入语音(最小化的误差)的码矢量来表示该子帧。

[0049] 编码激励108通常包括脉冲型信号或噪声型信号,这些在数学上构建或保存在码本中。该码本可用于编码器和接收方解码器。编码激励108,其可以是随机或固定码本,可以是(隐式或显式)硬编码到编解码器的矢量量化字典。此种固定码本可以是代数码激励线性预测或可以显式存储。

[0050] 码本中的码矢量乘以适当的增益调整以使能量等于输入语音的能量。相应地,编码激励108的输出在进入线性滤波器之前乘以增益G。107。

[0051] 短期线性预测滤波器103对码矢量的‘白’频谱进行整形以类似于输入语音的频谱。同样地,在时域中,短期线性预测滤波器103将短期相关性系数(与先前样本的相关性)并入白序列中。对激励进行整形的滤波器具有形式为 $1/A(z)$ 的全极点模型(短期线性预测滤波器103),其中 $A(z)$ 被称为预测滤波器并且可通过线性预测(例如,莱文森—德宾算法)获得。在一个或多个实施例中,可以使用全极点滤波器,因为它是人类声道的很好表现而且很容易计算。

[0052] 通过分析原始信号101获得短期线性预测滤波器103并且通过一组系数表示:

$$[0053] \quad A(z) = \sum_{i=1}^P 1 + a_i \cdot z^{-i}, i=1,2,\dots,P \quad (12)$$

[0054] 如先前所述,浊音语音的区域显示长期周期。这种周期,称为基音,被基音滤波器 $1/(B(z))$ 引入到合成频谱中。长期预测滤波器105的输出取决于基音和基音增益。在一个或多个实施例中,可以从原始信号、残差信号或加权原始信号中估计该基音。在一个实施例中,可以使用公式(13)表示长期预测函数 $B(z)$ 如下。

$$[0055] \quad B(z) = 1 - G_p \cdot z^{-\text{Pitch}} \quad (13)$$

[0056] 加权滤波器110与上述短期预测滤波器有关。可以如公式(14)所述表示其中一个典型的加权滤波器。

$$[0057] \quad W(z) = \frac{A(z/\alpha)}{1 - \beta \cdot z^{-1}} \quad (14)$$

[0058] 其中 $\beta < \alpha$, $0 < \beta < 1$, $0 < \alpha \leq 1$ 。

[0059] 在另一个实施例中,可以通过使用如下公式(15)中的一个实施例所示的带宽扩展从LPC滤波器中得出加权滤波器 $W(z)$ 。

$$[0060] \quad W(z) = \frac{A(z/\gamma_1)}{A(z/\gamma_2)} \quad (15)$$

[0061] 在公式(15)中, $\gamma_1 > \gamma_2$, 它们是极点向原点移动的因子。

[0062] 相应地,对于语音的每个帧,计算LPC和基音,并且更新滤波器。对于语音的每个子帧,选择产生‘最佳’滤波的输出的码矢量表示子帧。增益的对应量化值必须向解码器传输以便进行适当的解码。LPC和基音值还必须进行量化并且每帧发送以便重构解码器处的滤波器。相应地,向解码器传输编码激励索引、量化增益索引、量化长期预测参数索引和量化短期预测参数索引。

[0063] 图2示出了在实施本发明实施例中实施使用CELP解码器对原始语音进行解码期间执行的操作,如将下面描述的那样。

[0064] 通过将接收到的码矢量通过对应的滤波器在解码器处重构语音信号。因此,除了后处理之外的每个块具有如图1的编码器所述的相同定义。

[0065] 在接收方设备处接收和解开80已编码CELP比特流。对于每个接收到的子帧,使用接收到的编码激励索引、量化增益索引、量化长期预测参数索引和量化短期预测参数索引通过对应的解码器,例如,增益解码器81、长期预测解码器82和短期预测解码器83求出对应的参数。例如,可以从接收到的编码激励索引中确定激励脉冲的位置和幅度信号以及码激励402的代数码矢量。

[0066] 参考图2,解码器是若干块的组合,该解码器包括编码激励201、长期预测203、短期预测205。初始解码器还包括合成语音206之后的后处理块207。后处理还可包括短期后处理和长期后处理。

[0067] 图3示出了传统CELP编码器。

[0068] 图3示出了使用额外的自适应码本用于改进长期线性预测的基本CELP编码器。通过将自适应码本307和码激励308的贡献相加产生激励,码激励308可以是如先前所述的随机或固定码本。自适应码本中的条目包括激励的时延版本。这使得它有可能有效率地对周

期性信号,例如浊音,进行编码。

[0069] 参考图3,自适应码本307包括过去合成激励304或在基音周期内重复过去激励基音循环。当基音延迟很大或很长时,可以将它编码为整数值。当基音延迟很小或很短时,通常将它编码为更准确的分数值。使用基音的周期性信息以生成激励的自适应分量。这种激励分量随后由增益 G_p 305 (还称为基音增益) 调整。

[0070] 长期预测对于浊音语音编码非常重要,因为浊音语音具有强的周期。浊音语音的相邻基音周期彼此相似,这意味着在数学上,下面激励表达中的基音增益 G_p 很高或接近于1。所得的激励在公式(16)中可以表示为各个激励的组合。

$$[0071] \quad e(n) = G_p \cdot e_p(n) + G_c \cdot e_c(n) \quad (16)$$

[0072] 其中, $e_p(n)$ 是索引为 n 的样本序列的一个子帧,来自于自适应码本307,其包括经过反馈回路(图3)的过去激励304。 $e_p(n)$ 可以自适应地低通滤波为低频率区域,该低频率区域的周期和谐波通常比高频率区域的多。 $e_c(n)$ 来自编码激励码本308 (还称为固定码本),其是当前激励贡献。此外,例如通过使用高通滤波增强、基音增强、色散增强、共振峰增强和其它还可以增强 $e_c(n)$ 。

[0073] 对于浊音语音,自适应码本307中的 $e_p(n)$ 的贡献可能是主导的,而且基音增益 G_p 305的值约为1。通常更新每个子帧的激励。典型的帧大小是20毫秒,而典型的子帧大小是5毫秒。

[0074] 如图1所述,固定编码激励308在进入线性滤波器之间乘以增益 G_c 306。在通过短期线性预测滤波器303进行滤波之前将固定码本激励108和自适应码本307中的两个相乘的激励分量相加在一起。量化这两个增益(G_p 和 G_c)并向解码器传输。相应地,向接收方音频设备传输编码激励索引、自适应码本索引、量化增益索引和量化短期预测参数索引。

[0075] 在接收方设备处接收使用图3所示的设备进行编码的CELP比特流。图4示出了接收方设备的对应解码器。

[0076] 图4示出了对应于图3中的编码器的基本CELP解码器。图4包括接收来自主要解码器的合成语音407的后处理块408。该解码器类似于图3,除了自适应码本307。

[0077] 对于接收到的每个子帧,使用接收到的编码激励索引、量化编码激励增益索引、量化基音索引、量化自适应码本增益索引和量化短期预测参数索引以通过对应的解码器,例如,增益解码器81、基音解码器84、自适应码本增益解码器85和短期预测解码器83求出对应的参数。

[0078] 在各种实施例中,CELP解码器是若干块的组合并且包括编码激励402、自适应码本401、短期预测406和后处理器408。除了后处理,每个块具有如图3的编码器所述的相同定义。后处理还可包括短期后处理和长期后处理。

[0079] 如前所述,CELP主要用于通过受益于具体人类声音特征或人类嗓音发声模型对语音信号进行编码。为了更高效地对语音信号进行编码,可以将语音信号分类为不同类,而且以不同的方式对每类进行编码。浊音/清音分类或清音判决可能是所有不同类的所有分类中的一种重要和基本分类。对于每类,经常使用LPC或STP滤波器来表示频谱包络。但是对LPC滤波器的激励可能是不同的。清音信号可以利用噪声型激励进行编码。另一方面,浊音信号可以利用脉冲型激励进行编码。

[0080] 码激励块(参考图3的标号308和图4中的402)示出了固定码本(FCB)的位置以便进

行一般的CELP编码。从FCB中选择的码矢量由通常示为G_c306的增益调整。

[0081] 图5A和5B示出了利用带宽扩展(BWE)的编码/解码的示例。图5A示出了具有BWE边信息的编码器处的操作,而图5B示出了具有BWE的解码器处的操作。

[0082] 通过使用低带参数502对低带信号501进行编码。量化低带参数502,并且可以通过比特流信道503传输生成的量化索引。通过利用高带边参数505并使用少量比特对从音频/语音信号504中提取的高带信号进行编码。通过比特流信道506传输量化的高带边参数(边信息索引)。

[0083] 参考图5B,在解码器处,低带比特流507用于产生已解码低带信号508。高带边比特流510用于解码高带边参数511。在高带边参数511的帮助下从低带信号508生成高带信号512。通过组合低带信号508和高带信号512产生最终音频/语音信号509。

[0084] 图6A和6B示出了在没有传输边信息的情况下利用BWE的编码/解码的另一示例。图6A示出了编码器处的操作,而图6B示出了解码器处的操作。

[0085] 参考图6A,通过使用低带参数602对低带信号601进行编码。量化低带参数602以生成量化索引,可以通过比特流信道603传输该量化索引。

[0086] 参考图6B,在解码器处,低带比特流604用于产生已解码低带信号605。在没有传输边信息的情况下从低带信号605生成高带信号607。通过组合低带信号605和高带信号607产生最终音频/语音信号606。

[0087] 图7示出了使用CELP类型的编解码器时的浊音语音或谐波音乐的理想化激励频谱的示例。

[0088] 在移除LPC频谱包络之后,理想化激励频谱702几乎是平坦的。理想化的低带激励频谱701可以用作低带激励编码的参考。理想化的高带激励频谱703在解码器处不可获得。理论上,理想化或非量化的高带激励频谱的能量等级可以与低带激励频谱的几乎相同。

[0089] 实际上,合成或已解码激励频谱看起来并不如图7所示的理想化激励频谱那么好。

[0090] 图8示出了使用CELP类型的编解码器时的浊音语音或谐波音乐的已解码激励频谱的示例。

[0091] 在移除LPC频谱包络804之后,已解码激励频谱802几乎是平坦的。已解码低带激励频谱801在解码器处可获得。已解码低带激励频谱801的质量尤其在包络能量低的区域中变得更差或更失真。这是由于多个原因导致的。例如,两个主要原因在于:闭环CELP编码强调高能量区域比强调低能量区域的多,以及低频信号的波形匹配比高频信号的容易,因为高频信号变化更快。对于低比特率CELP编码,例如AMR-WB,通常不对高带进行编码,而是利用BWE技术在解码器中生成高带。在这种情况下,可以简单地从低带激励频谱801复制高带激励频谱803,以及可以从低带频谱能量包络预测或估计高带频谱能量包络。按照传统方式,6400Hz之后的生成的高带激励频谱803是从6400Hz之前的子带复制而来的。如果频谱质量从0Hz到6400Hz是等同的,这可能是个好方法。然而,对于低比特率CELP编解码器,频谱质量从0Hz到6400Hz可能差异较大。从6400Hz之前的低频带的末端区域复制而来的子带的质量可能较差,其随后将引入额外的噪声到6400Hz至8000Hz的高带区域中。

[0092] 扩展的高频带的带宽通常比已编码低频带的小得多。因此,在各种实施例中,选择低带中的最佳子带并且将其复制到高带区域中。

[0093] 高质量子带有可能存在于整个低频带内的任何位置处。高质量子带的最有可能的

位置是高频谱能量区域对应的区域,即频谱共振峰区域。

[0094] 图9示出了使用CELP类型的编解码器时的浊音语音或谐波音乐的已解码激励频谱的示例。

[0095] 在移除LPC频谱包络904之后,已解码激励频谱902几乎是平坦的。已解码低带激励频谱901在解码器处可获得,但是在高带903处不可获得。已解码低带激励频谱901的质量尤其在频谱包络904的能量较低的区域中变得更差或更失真。

[0096] 在图9的所示情况下,在一个实施例中,高质量子带位于第一语音共振峰区域周围(例如,在该示例实施例中约为2000Hz)。在各种实施例中,高质量子带可以位于0和6400Hz之间的任何位置处。

[0097] 在确定最佳子带的位置之后,如图9进一步所示,将其从低带内复制到高带。从而通过从所选择的子带复制以生成高带激励频谱903。图9中的高带903的感知质量因为改进的激励频谱听起来比图8中的高带803好得多。

[0098] 在一个或多个实施例中,如果低带频谱包络在频域中的解码器处可获得,则可以通过从所有子带候选中搜索最高子带能量确定最佳子带。

[0099] 可替代性地,在一个或多个实施例中,如果频域频谱包络不可获得,则还可以从反映频谱能量包络或频谱共振峰峰值的任何参数中确定高能量位置。BWE的最佳子带位置对应于最高频谱峰值位置。

[0100] 最佳子带起始点的搜索范围可取决于编解码器比特率。例如,对于非常低比特率的编解码器,搜索范围可以从0到 $6400-1600=4800\text{Hz}$ (2000Hz到4800Hz),假设高带的带宽是1600Hz。在另一示例中,对于中等比特率的编解码器,搜索范围可以从2000Hz到 $6400-1600=4800\text{Hz}$ (2000Hz到4800Hz),假设高带的带宽是1600Hz。

[0101] 由于频谱包络从一个帧到下一帧缓慢变化,所以最高频谱共振峰能量对应的最佳子带起始点通常变化缓慢。为了避免最佳子带起始点从一个帧到另一个帧发生波动或频繁变化,可以在时域中的相同浊音区域内采用一些平滑处理,除非频谱峰值能量从一个帧到下一个帧发生巨大变化或者产生新的浊音区。

[0102] 图10示出了根据用于实施子带位移或复制BWE的本发明实施例的解码器处的操作。

[0103] 通过使用接收到的比特流1001对时域低带信号1002进行解码。低带时域激励1003通常在解码器处可获得。有时,低带频域激励也可获得。如果不可获得,可以将低带时域激励1003变换到频域以获得低带频域激励。

[0104] 浊音语音或音乐信号的频谱包络通常通过LPC参数表示。有时,直接频域频谱包络在解码器处可获得。在任何情况下,能量分布信息1004可以从LPC参数或从直接频域频谱包络或DFT域或FFT域等任何参数中提取。通过使用低带能量分布信息1004,最佳子带通过搜索相对较高的能量峰值从低带中选择。然后从低带复制所选择的子带到高带区域。随后将预测或估计的高带频谱包络应用到高带区域,或者时域高带激励1005经过表示高带频域包络的预测或估计的高带滤波器。高带滤波器的输出是高带信号1006。通过组合低带信号1002和高带信号1006获得最终的语音/音频输出信号1007。

[0105] 图11示出了用于实施子带位移或复制BWE的解码器的替代性实施例。

[0106] 与图10不同,图11假设频域低带频谱可获得。通过简单地搜索频域中的相对较高

的能量峰值选择低频带中的最佳子带。随后,从低带复制所选择的子带到高带。在应用估计的高带频谱包络之后,形成高带频谱1103。通过组合低带频谱1102和高带频谱1103获得最终频域语音/音频频谱。通过将频域/语音/音频频谱转换到时域产生最终时域语音/音频信号输出。

[0107] 当滤波器组分析和合成在包含所需频谱范围的解码器处可获得,SBR算法可以通过从滤波器组分析复制对应于所选择的低带的输出的低频带系数到高频带区域实现频带位移。

[0108] 图12示出了根据本发明实施例的在解码器处执行的操作。

[0109] 参考图12,一种在解码器处解码已编码音频比特流的方法包括接收已编码音频比特流。在一个或多个实施例中,接收到的音频比特流已经进行了CELP编码。尤其是,通过CELP只对低频带进行编码。CELP在较高频谱能量区域中产生的频谱质量比在较低频谱能量区域中产生的相对较高。相应地,本发明实施例包括解码音频比特流以生成已解码低带音频信号和对应于低频带的低带激励频谱(方框1210)。使用已解码低带音频信号的频谱包络的能量信息从低频带内选择子带区域(方框1220)。通过从所选择的子带区域复制子带激励频谱到对应于高频带的高子带区域生成高频带的高带激励频谱(方框1230)。使用高带激励频谱生成音频输出信号(方框1240)。尤其是,使用生成的高带激励频谱通过应用高带频谱包络生成扩展的高带音频信号。将扩展的高带音频信号添加到已解码低带音频信号以生成具有扩展的频率带宽的音频输出信号。

[0110] 如使用图10和11先前所述,本发明实施例可通过不同方式应用,取决于频域频谱包络是否可获得。例如,如果频域频谱包络可获得,则可以选择具有最高子带能量的子带。另一方面,如果频域频谱包络不可获得,则频谱包络的能量分布可以从线性预测编码(LPC)参数、离散傅里叶变换(DFT)域,或快速傅里叶变换(FFT)域参数确定。类似地,如果频谱共振峰峰值信息可获得(或可计算的),则可以在一些实施例中使用。如果只有低带时域激励可获得,则可以通过将低带时域激励变换到频域计算低带频域激励。

[0111] 在各种实施例中,可以使用本领域普通技术人员已知的任何已知方法计算频谱包络。例如,在频域中,频谱包络可以是简单的一组能量,表示一组子带的能量。类似地,在另一示例中,频谱包络在时域中可以通过LPC参数表示。LPC参数在各种实施例中可能具有许多形式,例如反射系数、LPC系数、LSP系数、LSF系数。

[0112] 图13A和13B示出了根据本发明实施例的实施带宽扩展的解码器。

[0113] 参考图13A,用于解码已编码音频比特流的解码器包括低带解码单元1310,用于解码音频比特率以生成对于低频带的低带激励频谱。

[0114] 解码器还包括带宽扩展单元1320,其耦合到低带解码单元1310并且包括子带选择单元1330和复制单元1340。子带选择单元1330用于使用已解码音频比特流的频谱包络的能量信息从低频带内选择子带区域。复制单元1340用于通过从所选择的子带区域复制子带激励频谱到对应于高频带的高子带区域生成高频带的高带激励频谱。

[0115] 高带信号生成器1350耦合到复制单元1340。高带信号生成器1350用于使用预测的高带频谱包络生成高带时域信号。输出生成器耦合到高带信号生成器1350和低带解码单元1310。输出生成器1360用于通过组合解码音频比特流获得的低带时域信号和高带时域信号生成音频输出信号。

[0116] 图13B示出了实施带宽扩展的解码器的替代性实施例。

[0117] 类似于图13A,图13B的解码器还包括低带解码单元1310和带宽扩展单元1320,带宽扩展单元1320耦合到低带解码单元1310并且包括子带选择单元1330和复制单元1340。

[0118] 参考图13B,解码器还包括高带频谱生成器,其耦合到复制单元1340。高带信号生成器1355用于使用高带频谱包络能量以通过高带激励频谱生成高频带的高带频谱。

[0119] 输出频谱生成器1365耦合到高带频谱生成器1355和低带解码单元1310。输出频谱生成器用于通过组合解码来自低带解码单元1310的音频比特流获得的低带频谱和来自高带频谱生成器1355的高带频谱生成频域音频频谱。

[0120] 反变换信号生成器1370用于通过将频域音频频谱反变换到时域生成时域音频信号。

[0121] 图13A和13B所述的各种部件可以在一个或多个实施例中的硬件中实施。在一些实施例中,它们在软件中实施并且用于在信号处理器中运作。

[0122] 相应地,本发明实施例可用于改进解码CELP编码的音频比特流的解码器处的带宽扩展。

[0123] 图14示出了根据本发明实施例的通信系统10。

[0124] 通信系统10具有经由通信链路38和40耦合到网络36的音频接入设备7和8。在一个实施例中,音频接入设备7和8是基于IP的语音传输(VOIP)设备,以及网络36是广域网(WAN)、公共交换电话网络(PSTN)和/或互联网。在另一实施例中,通信链路38和40是有线和/或无线宽带连接。在另一替代性实施例中,音频接入设备7和8是蜂窝或移动电话,链路38和40是无线移动电话信道,而且网络36表示移动电话网络。

[0125] 音频接入设备7使用麦克风12将声音,例如音乐或人的声音转换到模拟音频输入信号28。麦克风接口16将模拟音频输入信号28转换成数字音频信号33以输入到编解码器20的编码器22中。根据本发明实施例,编码器22产生已编码音频信号TX以便经由网络接口26向网络26传输。编解码器20内的解码器24经由网络接口26接收来自网络36的已编码音频信号RX,并且将已编码音频信号RX转换成数字音频信号34。扬声器接口18将数字音频信号34转换成适用于驱动扬声器14的音频信号30。

[0126] 在本发明实施例中,当音频接入设备7为VOIP设备时,音频接入设备7内的一些或所有分量在手机内实施。然而,在一些实施例中,麦克风12和扬声器14为单独的单元,而且麦克风接口16、扬声器接口18、编解码器20和网络接口26在个人计算机内实施。编解码器20可以在运行在计算机或专用处理器上的软件中实施或通过,例如专用集成电路(ASIC)上的专用硬件实施。麦克风接口16通过模数(A/D)转换器,以及位于手机和/或计算机内的其它接口电路实施。同样地,扬声器接口18通过数模转换器和位于手机和/或计算机内的其它接口电路实施。在其它实施例中,音频接入设备7可以通过本领域已知的其它方式实施和划分。

[0127] 在本发明实施例中,当音频接入设备7为蜂窝或移动电话时,音频接入设备7内的元件在蜂窝手机内实施。编解码器20通过运行在手机内的处理器上的软件或通过专用硬件实施。在本发明的其它实施例中,音频接入设备可以在诸如端到端有线和无线数字通信系统,例如对讲机和无线手机,之类的其它设备中实施。在客户音频设备等应用中,音频接入设备可包括数字麦克风系统或音乐播放设备中的只具有例如编码器22或解码器24的编解

码器。在本发明的其它实施例中,编解码器20可以在没有麦克风12和扬声器14的情况下在接入PSTN的蜂窝基站中使用。

[0128] 本发明各种实施例中描述的用于改进清音/浊音分类的语音处理可以在,例如,编码器22或解码器24中实施。用于改进清音/浊音分类的语音处理可以在各种实施例中的硬件或软件中实施。例如,编码器22或解码器24可以是数字信号处理(DSP)芯片的一部分。

[0129] 图15示出了处理系统的方框图,该处理系统可以用来实现本文公开的设备和方法。特定设备可利用所有所示的组件或所述组件的仅一子集,且设备之间的集成程度可能不同。此外,设备可以包括部件的多个实例,例如多个处理单元、处理器、存储器、发射器、接收器等。处理系统可以包括配备一个或多个输入/输出设备,例如扬声器、麦克风、鼠标、触摸屏、按键、键盘、打印机、显示器等的处理单元。处理单元可以包括中央处理器(CPU)、存储器、大容量存储器设备、视频适配器以及连接至总线的I/O接口。

[0130] 总线可以是任意类型的若干总线架构中的一个或多个,包括存储总线或存储控制器、外设总线、视频总线等等。CPU可包括任意类型的电子数据处理器。存储器可包括任何类型的系统存储器,例如静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、同步DRAM(SDRAM)、只读存储器(ROM)或其组合等等。在实施例中,存储器可包括在开机时使用的ROM以及执行程序时使用的程序和数据存储器的DRAM。

[0131] 大容量存储器设备可包括任何类型的存储器设备,其用于存储数据、程序和其它信息,并使这些数据、程序和其它信息通过总线访问。大容量存储器设备可包括如下项中的一项或多项:固态磁盘、硬盘驱动器、磁盘驱动器、光盘驱动器等等。

[0132] 显示卡以及I/O接口提供接口以将外部输入以及输出设备耦合到处理单元上。如图所示,输入以及输出设备的实例包含耦合到显示卡上的显示器以及耦合到I/O接口上的鼠标/键盘/打印机。其它设备可以耦合到处理单元上,并且可以利用额外的或较少的接口卡。例如,可使用如通用串行总线(USB)(未示出)等串行接口将接口提供给打印机。

[0133] 处理单元还包含一个或多个网络接口,所述网络接口可以包括例如以太网电缆或其类似者等有线链路,和/或用以接入节点或不同网络的无线链路。网络接口允许处理单元经由网络与远程单元通信。举例来说,网络接口可以经由一个或多个发射器/发射天线以及一个或多个接收器/接收天线提供无线通信。在一个实施例中,处理单元耦合到局域网或广域网上以用于数据处理以及与远程设备通信,所述远程设备例如其它处理单元、因特网、远程存储设施或其类似者。

[0134] 虽然已参考说明性实施例描述了本发明,但此描述并不意图限制本发明。所属领域的技术人员在参考该描述后,将会明白说明性实施例的各种修改和组合,以及本发明其他实施例。例如,上述各种实施例可以彼此组合。

[0135] 虽然已详细地描述了本发明及其优点,但是应理解,可以在不脱离如所附权利要求书所界定的本发明的精神和范围的情况下对本发明做出各种改变、替代和更改。例如,上文所论述的许多特征和功能可以由软件、硬件、固件或其组合来实施。此外,本发明的范围并不局限于说明书中所述的过程、机器、制造、物质组分、构件、方法和步骤的具体实施例。所属领域的一般技术人员可从本发明中轻易地了解,可根据本发明使用现有的或即将开发出的,具有与本文所描述的相应实施例实质相同的功能,或能够取得与所述实施例实质相同的结果的流程、机器、制造、物质组分、构件、方法或步骤。相应地,所附权利要求范围包括

这些流程、机器、制造、物质组分、构件,方法,及步骤。

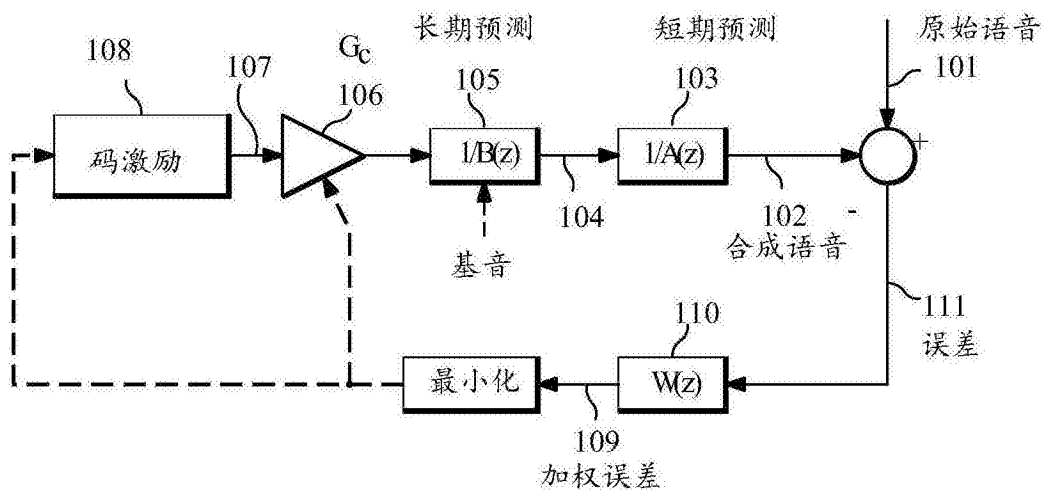


图1

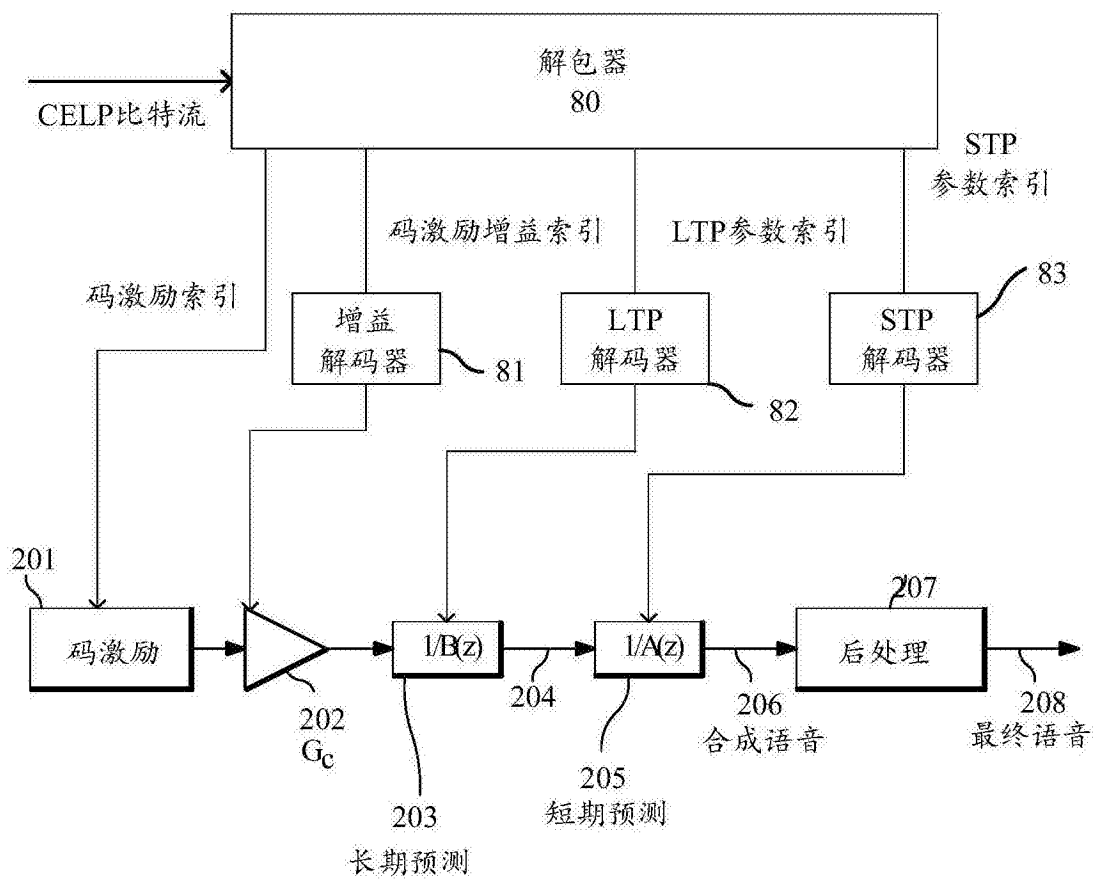


图2

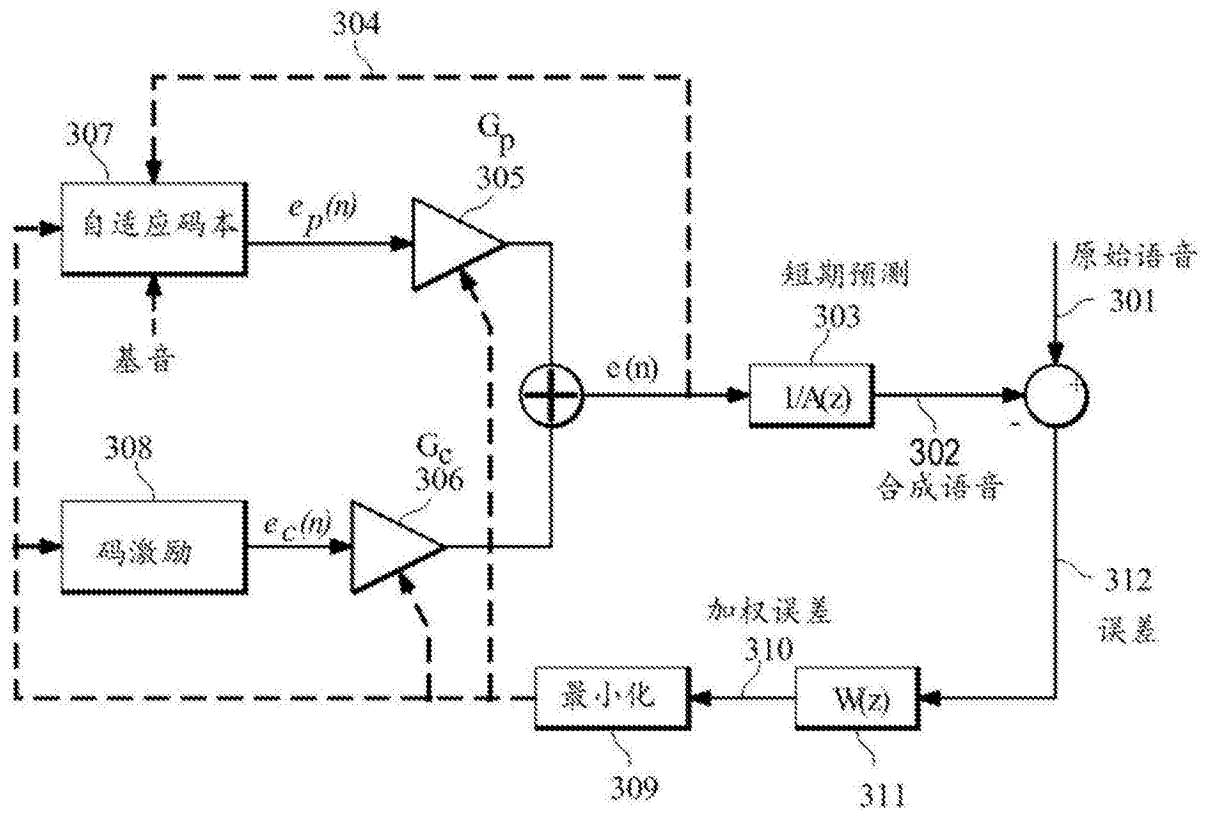


图3

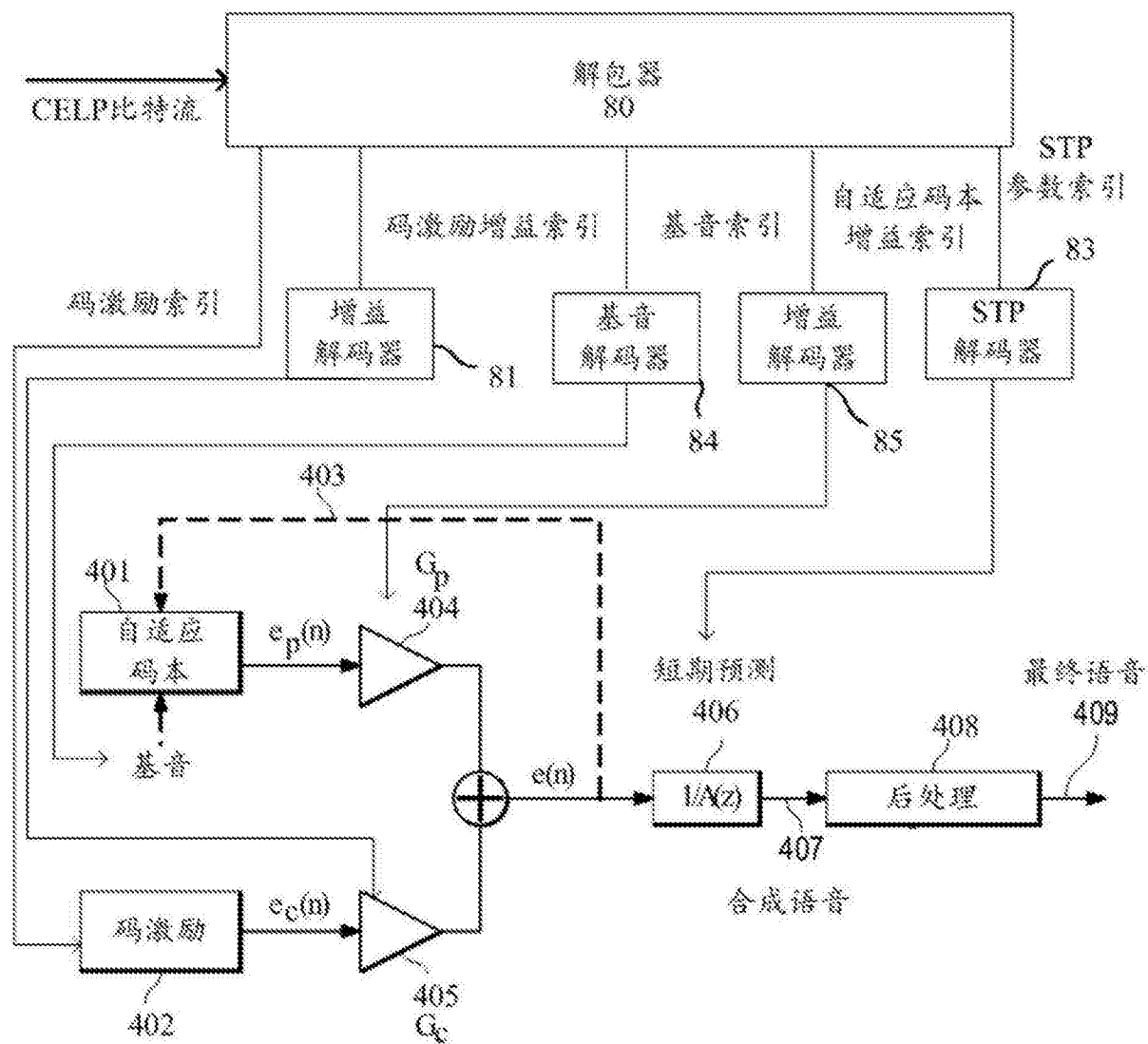


图4

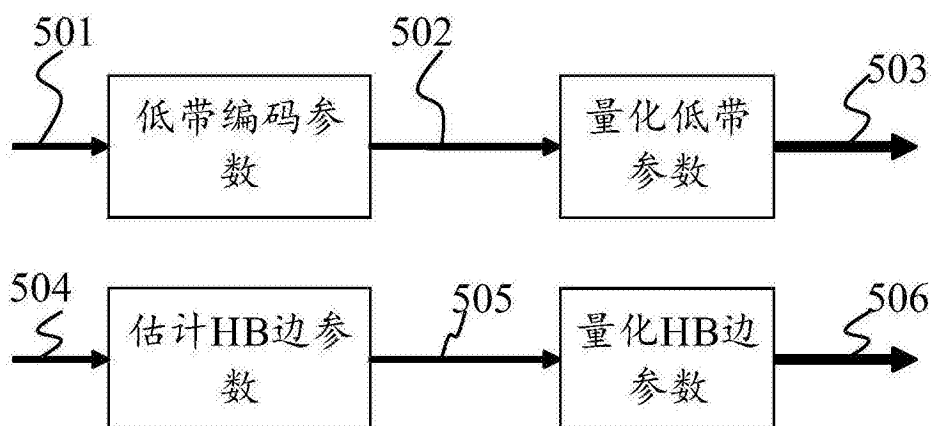


图5A

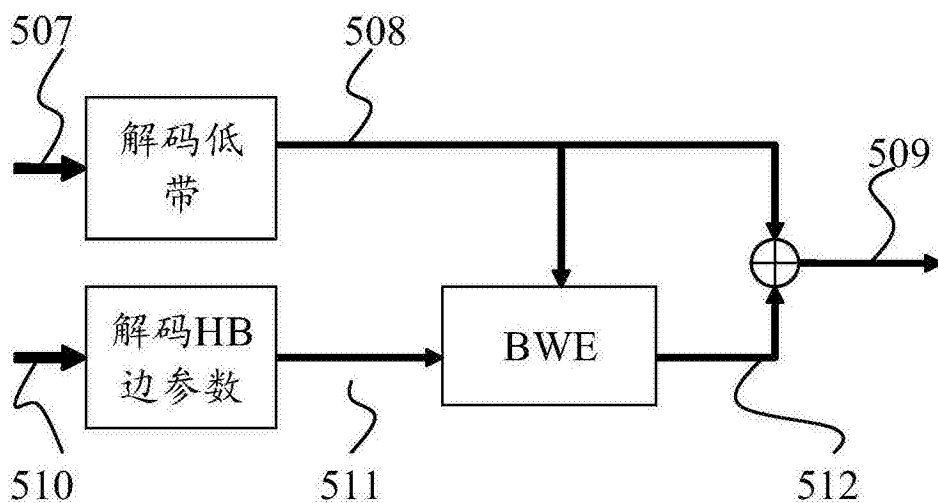


图5B

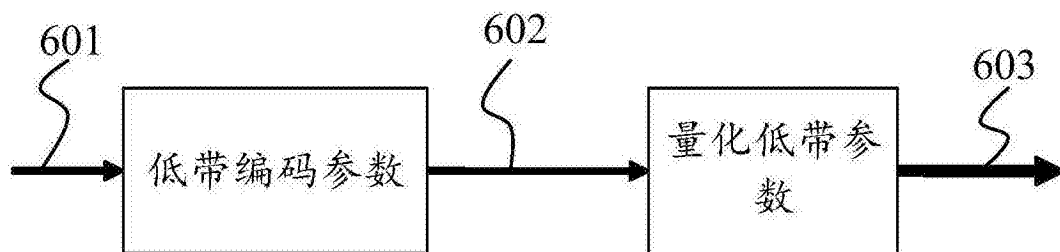


图6A

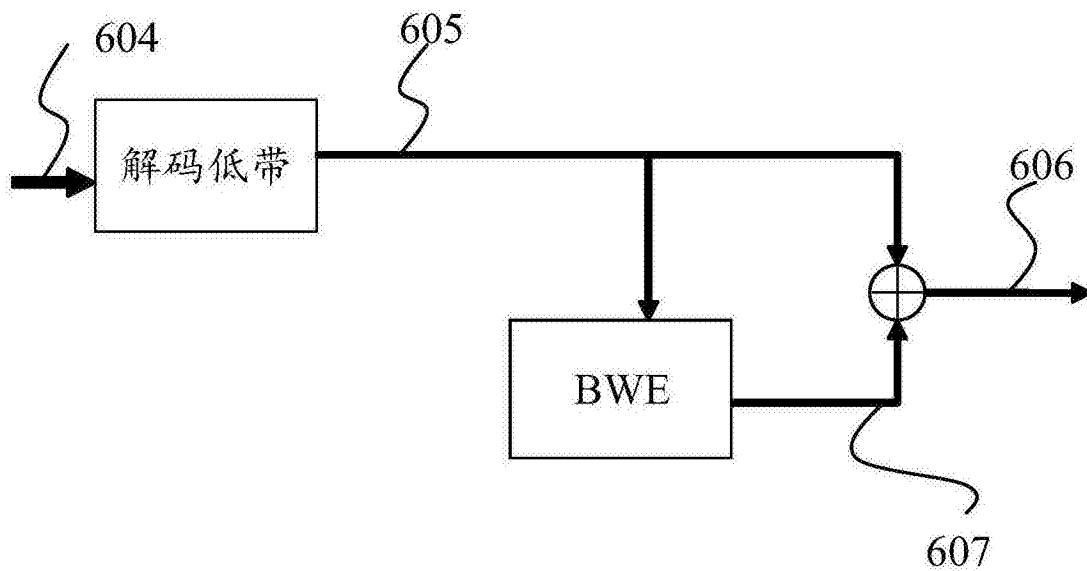


图6B

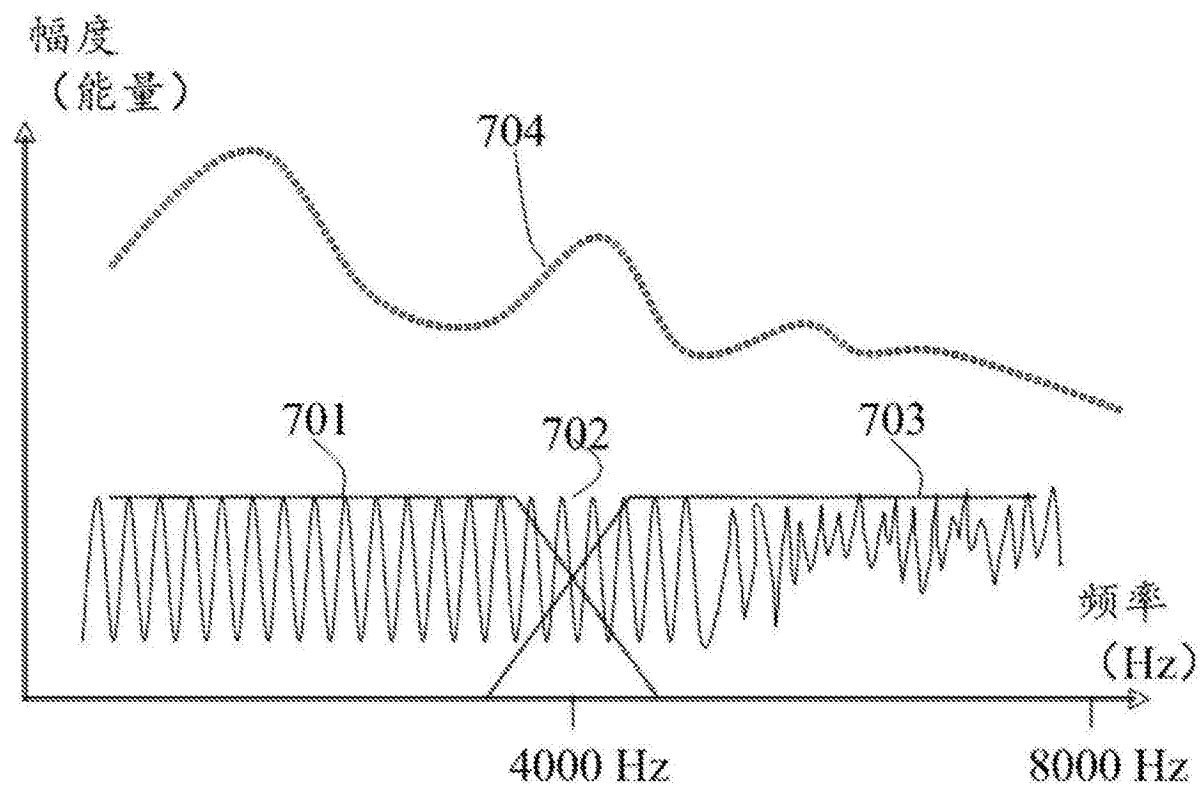


图7

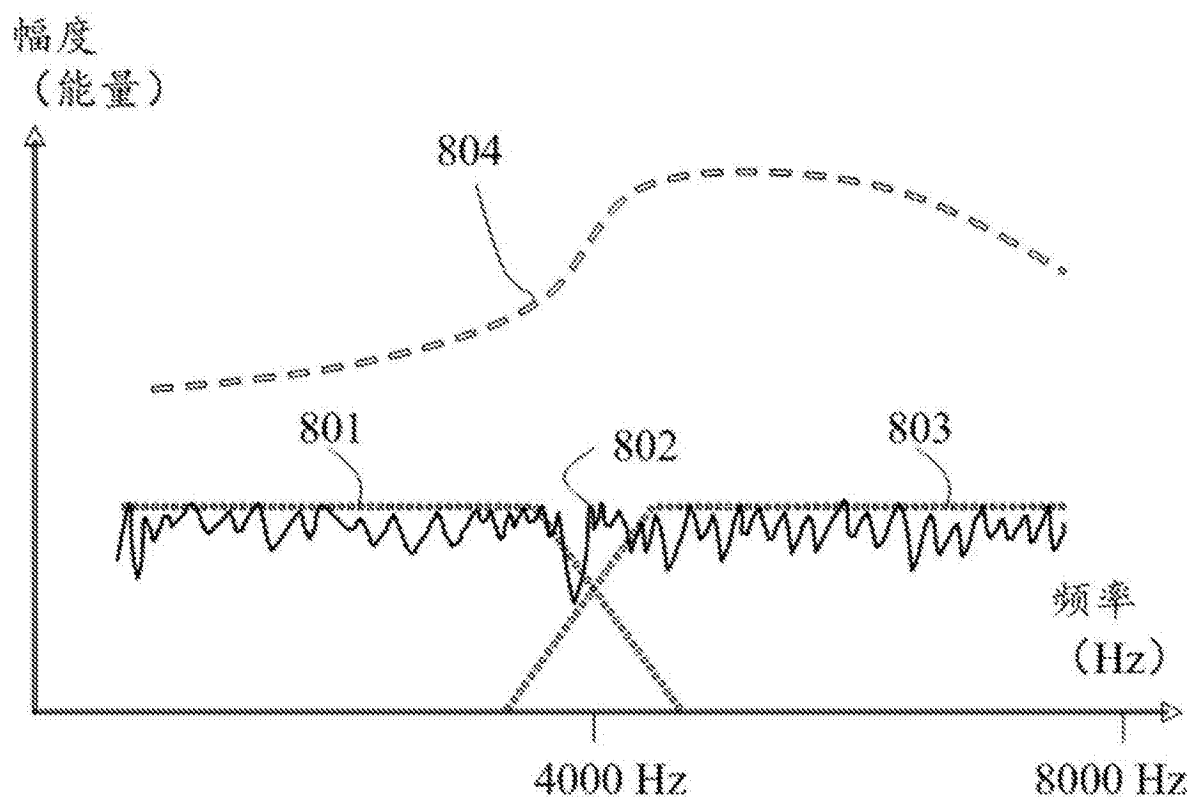


图8

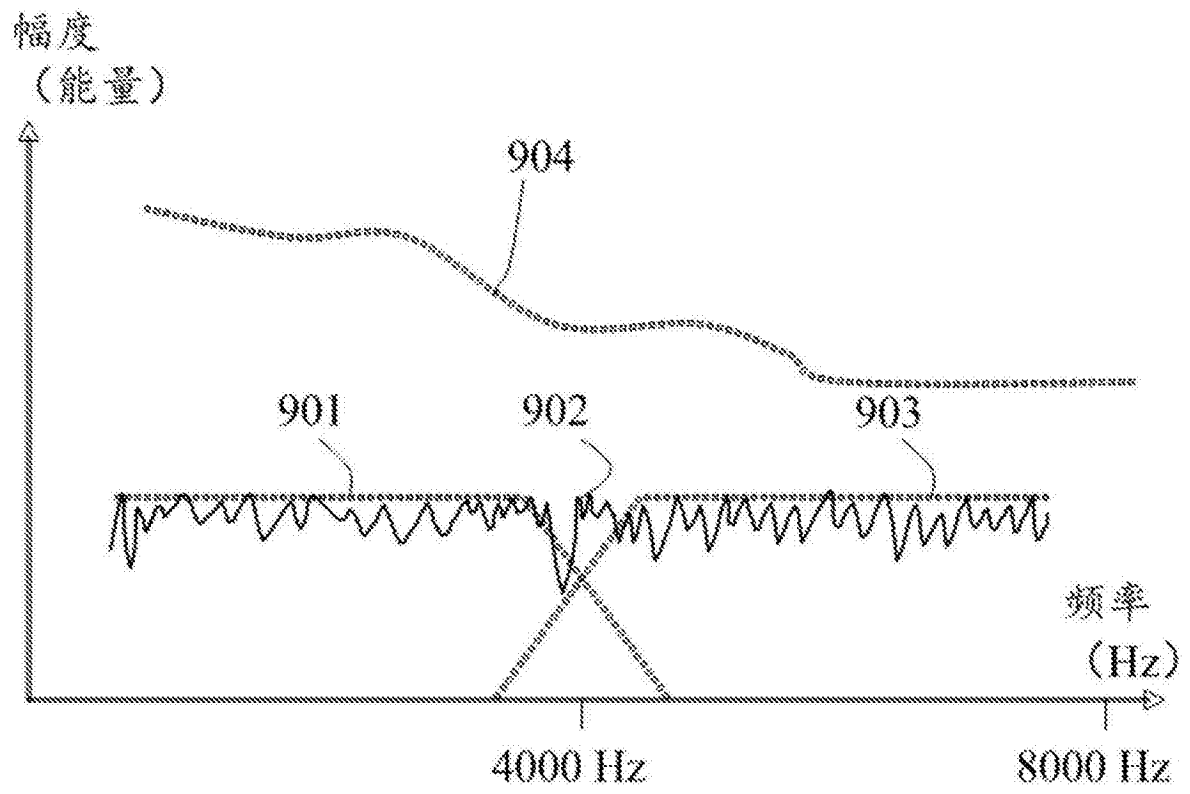


图9

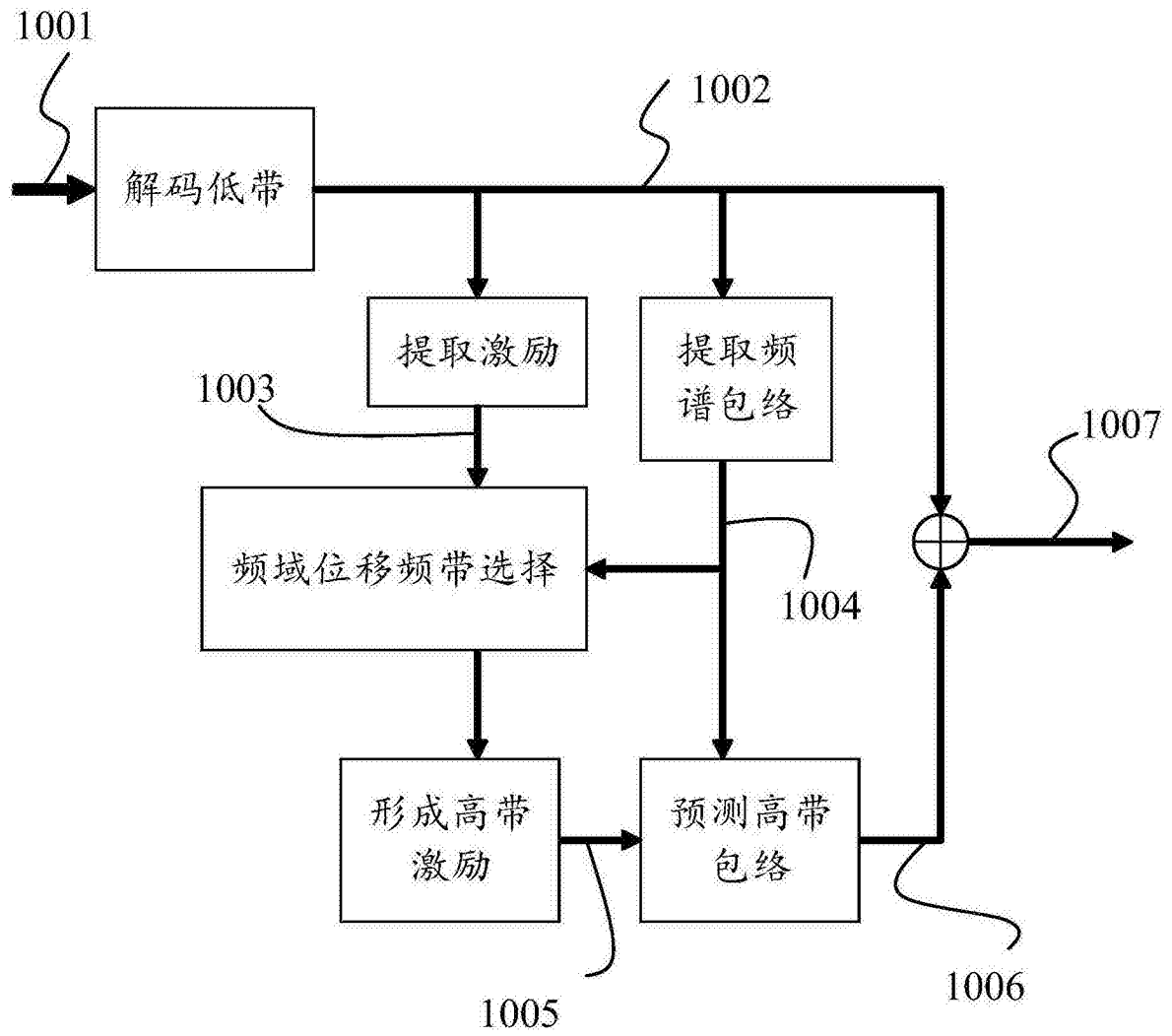


图10

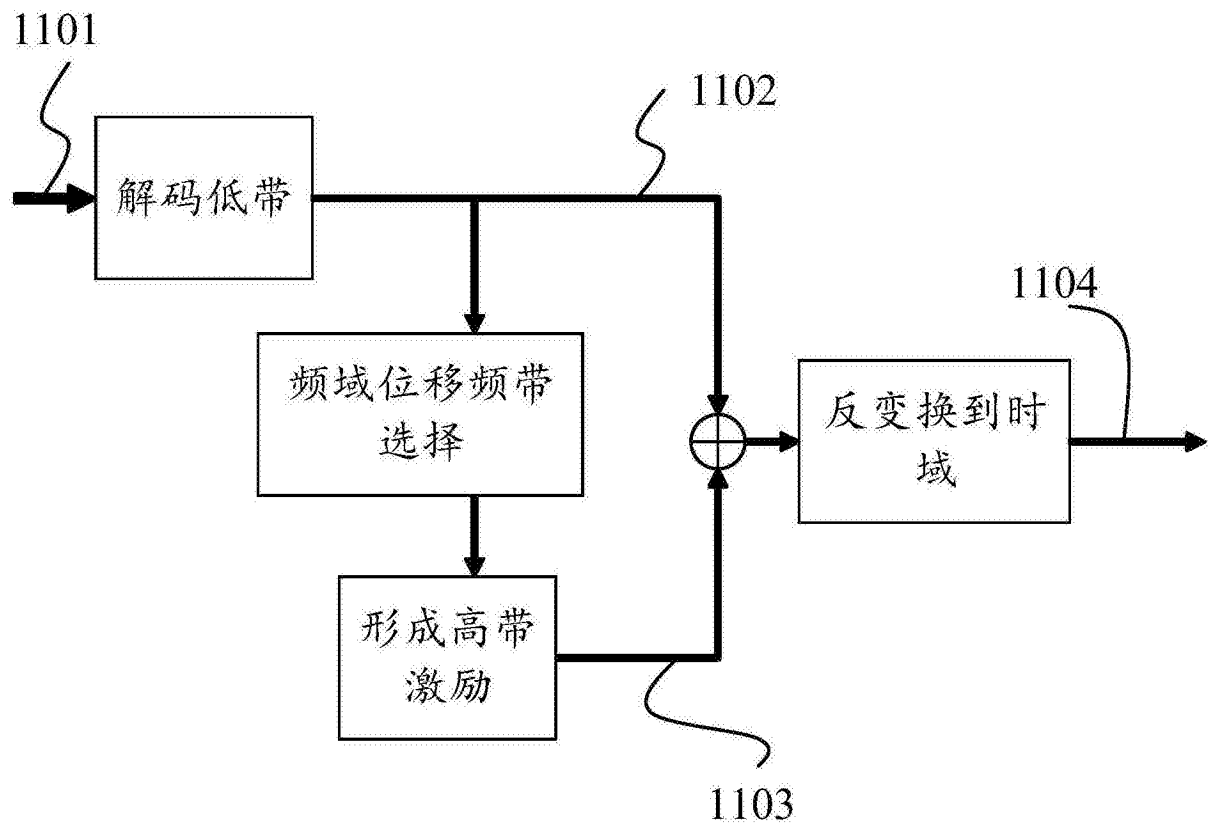


图11

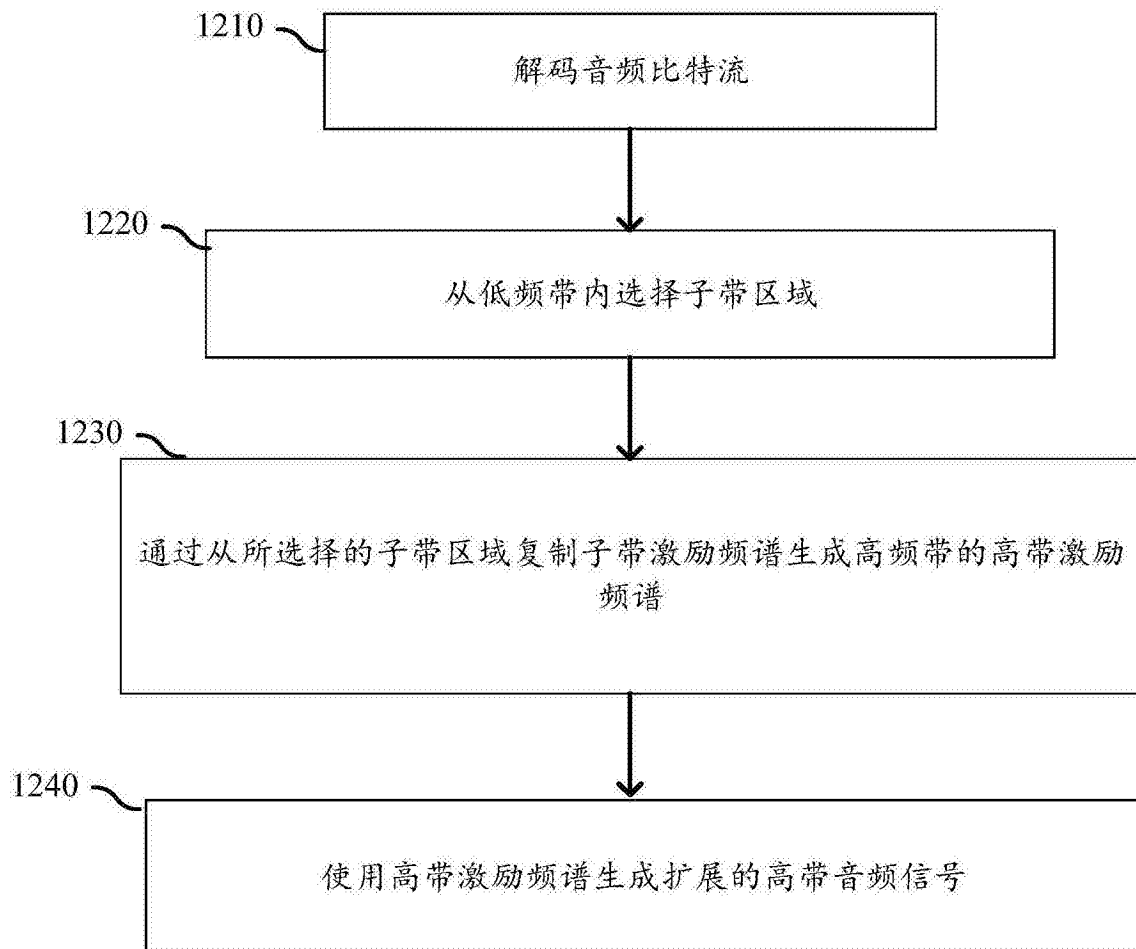


图12

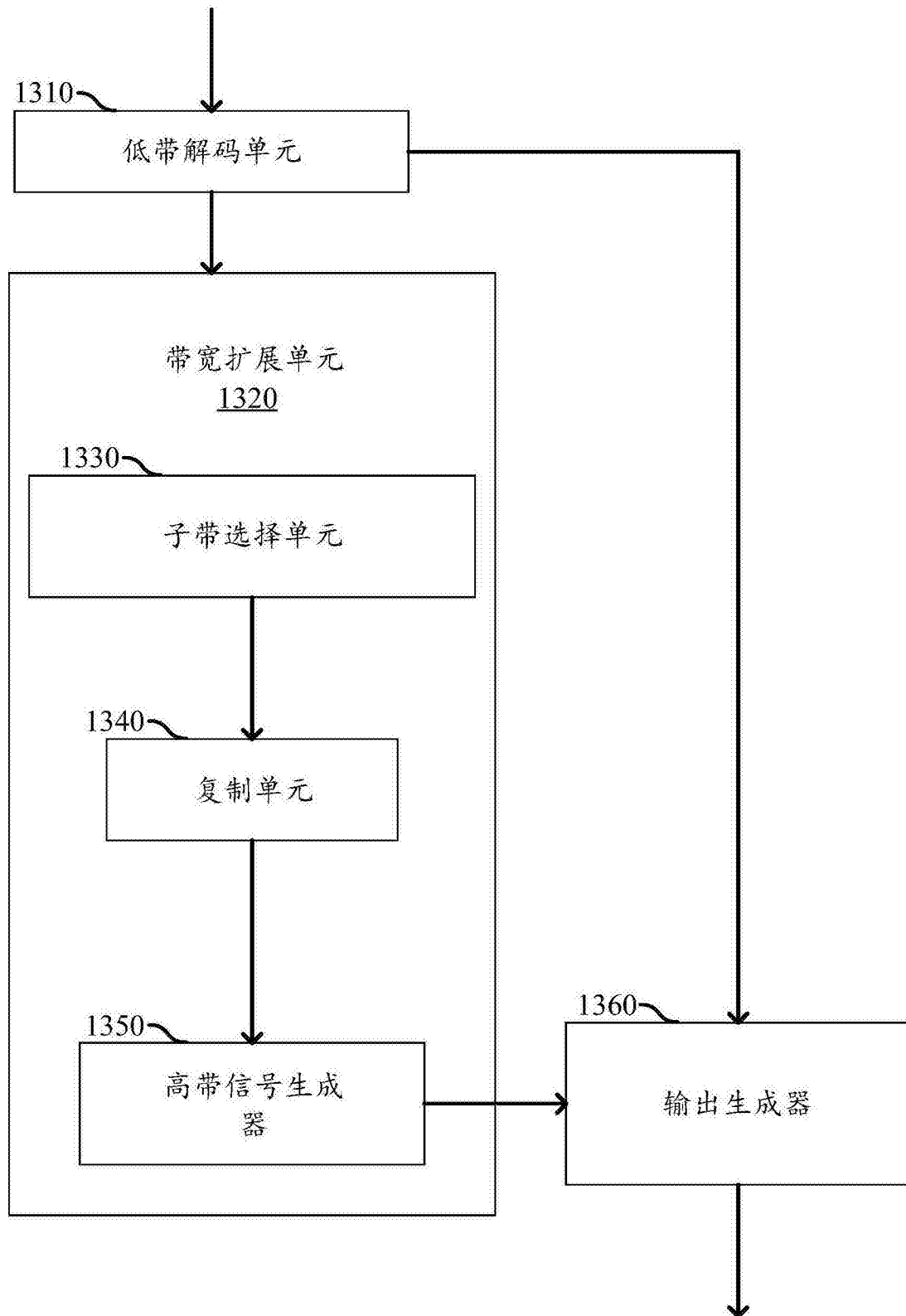


图13A

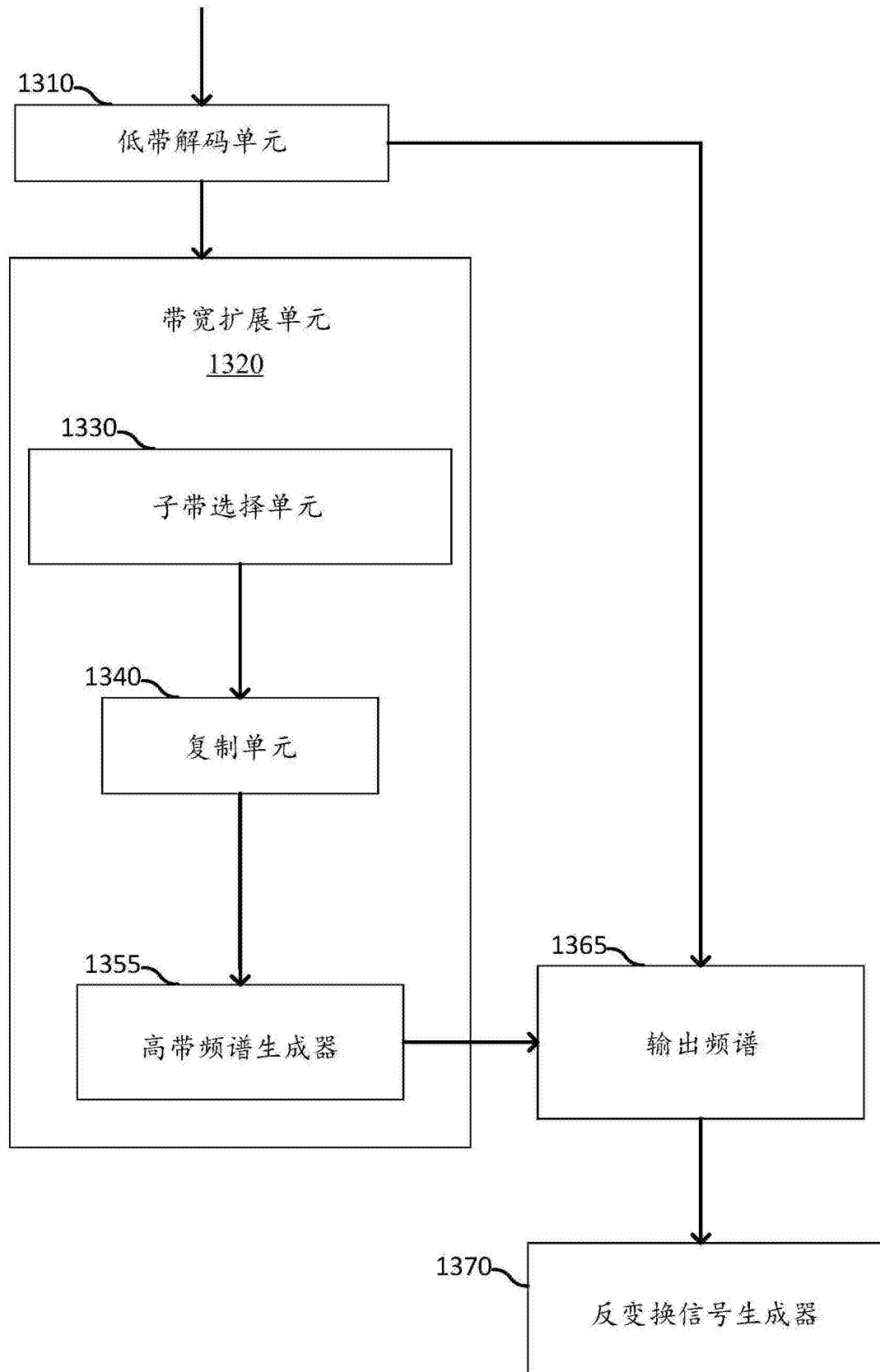


图13B

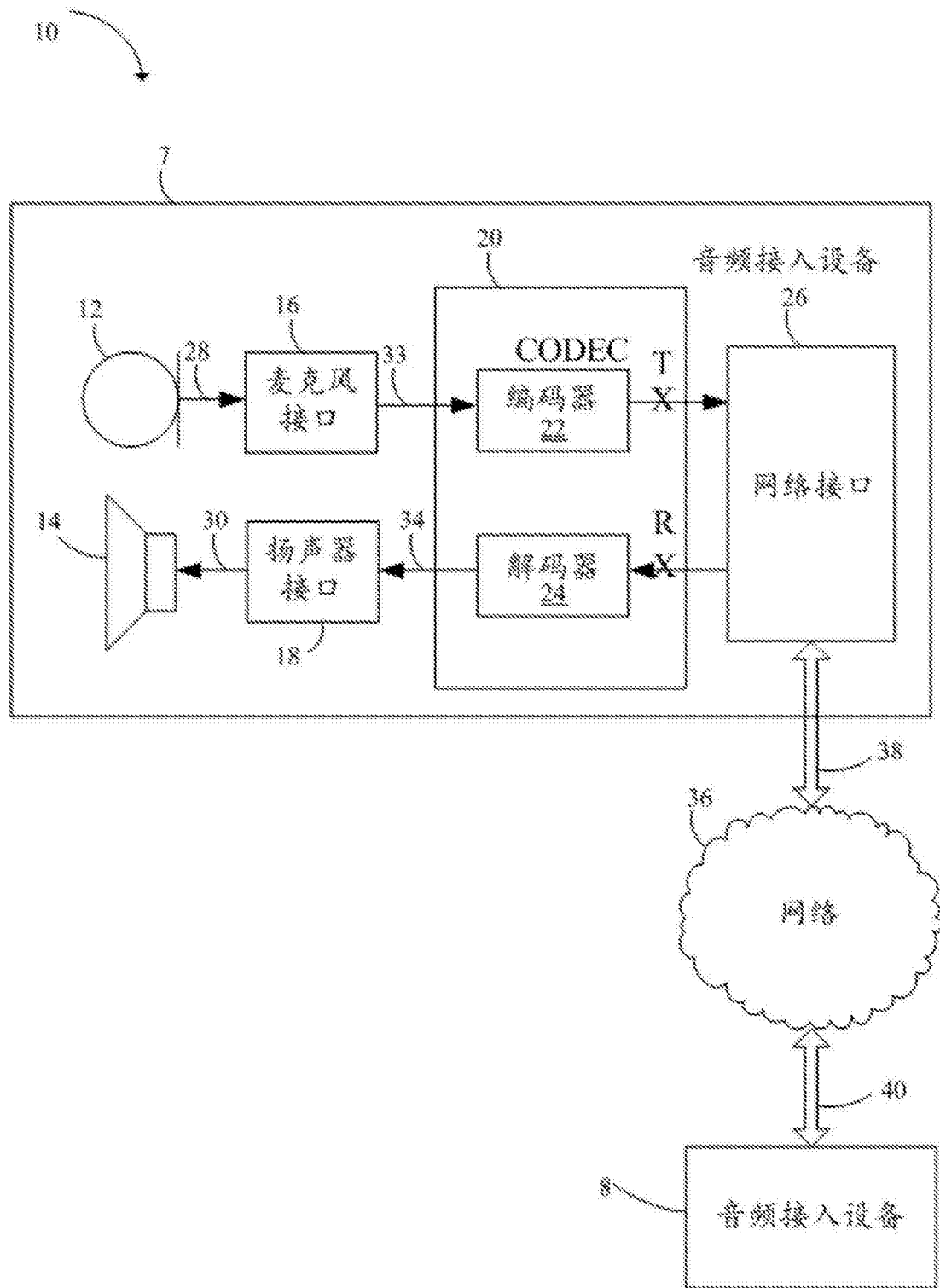


图14

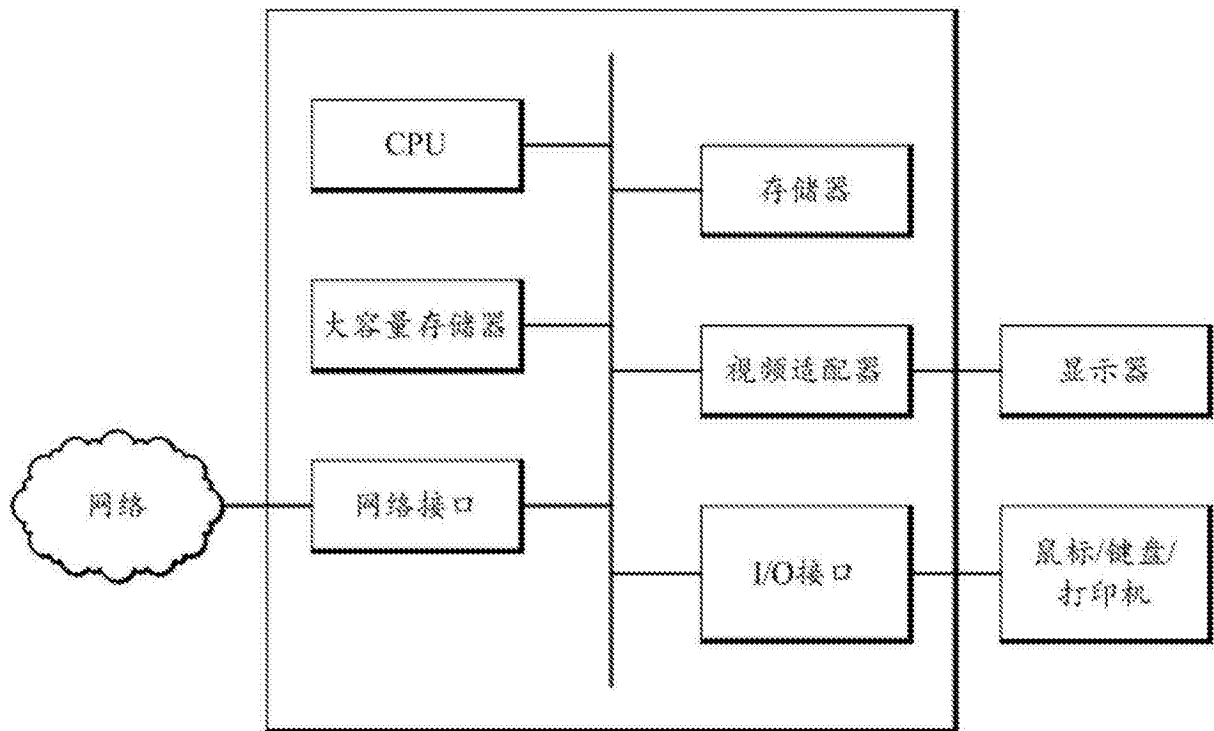


图15