

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200780051962.8

[51] Int. Cl.

G10L 19/14 (2006.01)

G10L 19/08 (2006.01)

G10L 19/04 (2006.01)

G10L 19/12 (2006.01)

G10L 21/02 (2006.01)

[43] 公开日 2010年1月6日

[11] 公开号 CN 101622666A

[22] 申请日 2007.12.14

[21] 申请号 200780051962.8

[30] 优先权

[32] 2007.3.2 [33] US [31] 60/892,667

[86] 国际申请 PCT/SE2007/051000 2007.12.14

[87] 国际公布 WO2008/108702 英 2008.9.12

[85] 进入国家阶段日期 2009.9.2

[71] 申请人 艾利森电话股份有限公司

地址 瑞典斯德哥尔摩

[72] 发明人 S·布鲁恩

[74] 专利代理机构 中国专利代理(香港)有限公司

代理人 王岳 李家麟

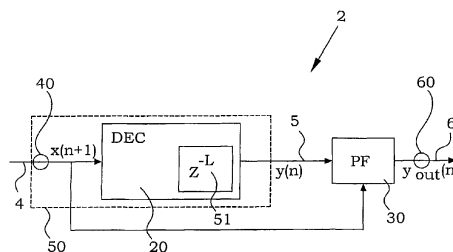
权利要求书 3 页 说明书 16 页 附图 9 页

[54] 发明名称

非因果后置滤波器

[57] 摘要

一种解码器装置包括用于基于帧的编码信号参数(4)的接收器输入端(40)，以及被设置为基于参数(4)提供解码音频信号(5)的帧的解码器(20)。接收器输入端(40)和/或解码器(20)被设置为建立在接收器输入端(40)可得到第一帧的参数的时刻与在解码器(20)的输出端可得到该第一帧的解码音频信号的帧的时刻之间的时间差，该时间差对应于至少一帧。后置滤波器(30)被连接至解码器(20)的输出端以及接收器输入端(40)。后置滤波器(30)被设置为响应于各个后续帧的参数(4)，提供对解码音频信号的帧的滤波以成为输出信号(6)。



1、一种解码器电路装置，包括：

接收器输入端（40），用于基于帧的编码信号参数（4）；

解码器（20），被连接至所述接收器输入端（40），该解码器被设置为基于所述参数提供解码音频信号（5；54）的帧；

后置滤波器（30；30A，30B），被连接至所述解码器（20）的输出端并被设置为基于解码音频信号（5；54）的所述帧提供输出信号；以及

输出端（60），用于所述输出信号（6），

其特征在于，

所述接收器输入端（40）和所述解码器（20）中至少之一被设置为建立在所述接收器输入端（40）可得到第一帧的参数时刻与在所述解码器（20）的所述输出端可得到所述第一帧的解码音频信号时刻之间的时间差，该时间差对应于至少一帧；

所述后置滤波器（30；30A，30B）被连接至所述接收器输入端（40）；并且

所述后置滤波器（30；30A，30B）被设置为响应于各个后续帧的所述参数（4），提供对解码音频信号（5；54）的所述帧的滤波以成为输出信号（6）。

2、根据权利要求1所述的解码器电路装置，其中所述接收器输入端（40）包括用于至少两个相邻帧的参数的存储器（41），由此所述解码器（20）被提供以第一帧的参数（4A）并且所述后置滤波器（30；30A，30B）有权访问后续第二帧的参数（4B）。

3、根据权利要求1所述的解码器电路装置，其中所述解码器（20）包括在解码音频信号的所述帧被输出至所述后置滤波器（30；30A，30B）之前延迟该解码音频信号的所述帧的装置（51；53）。

4、根据权利要求1至3中任一项所述的解码器电路装置，其中所述后置滤波器（30；30A，30B）包括音调后置滤波器，由此所述音调后置滤波器中所使用的音调参数是基于所述后续帧的音调参数的。

5、根据权利要求4所述的解码器电路装置，其中所述后置滤波器（30；30A，30B）的所述音调后置滤波器被设置为用于为随后的子帧确定由所述随后的子帧的音调值减小的时间索引值；并且如果所述确定的

值大于或等于当前时间索引，则将所述随后的子帧的所述音调值作为所述当前帧的音调引导参数。

6、根据权利要求4或5所述的解码器电路装置，包括音频特性检测器，该音频特性检测器的输出端被连接至所述后置滤波器（30；30A，30B）；

所述后置滤波器（30；30A，30B）被设置为：根据先前帧和后续帧中的至少一个与当前帧之间的音频特性变化的显著性的测量，来调节所述音调后置滤波器中所使用的所述音调参数基于所述后续帧的所述音调参数的程度。

7、根据权利要求6所述的解码器电路装置，其中所述音频特性检测器是语音活动检测器和发声检测器中的至少一个，并且其中所述后置滤波器被设置为：在所检测的浊音语音起始时，将所述音调后置滤波器中所使用的所述音调参数以所述后续帧的所述音调参数为基础。

8、根据权利要求1至7中任一项所述的解码器电路装置，其中所述后置滤波器（30；30A，30B）被设置为还有权访问所述后续帧的解码信号。

9、根据权利要求1至8中任一项所述的解码器电路装置，其中所述解码器（20）是可伸缩解码器（120）或可伸缩解码器的部分，其中所述可伸缩解码器的次解码器（25）具有比所述可伸缩解码器的主解码器（21）高的时延。

10、一种解码器电路装置，包括可伸缩解码器（120）和至少两个根据权利要求7所述的解码器电路装置。

11、一种解码方法，包括以下步骤：

接收（210）基于帧的编码信号参数；

将所述参数解码（212）成解码音频信号的帧；

所述接收步骤和所述解码步骤中的至少之一引起在接收后可得到第一帧的参数的时刻与在解码后可得到该第一帧的解码音频信号的时刻之间的时间差，该时间差对应于至少一帧；

响应于各个后续帧的所述参数，将解码音频信号的所述帧后置滤波（214）成输出信号；以及

输出（216）所述输出信号。

12、根据权利要求11所述的解码方法，包括以下步骤：

在每个时刻存储至少两个相邻帧的参数，由此利用第一帧的参数执行所述解码步骤，并且利用对后续第二帧的参数的访问来执行所述后置滤波。

13、根据权利要求 11 所述的解码方法，包括以下步骤：

在执行所述后置滤波的步骤之前延迟解码音频信号的所述帧。

14、根据权利要求 11 至 13 中任一项所述的解码方法，其中所述后置滤波（214）步骤包括音调后置滤波，由此所述音调后置滤波中所使用的音调参数基于所述后续帧的音调参数。

15、根据权利要求 14 所述的解码方法，其中所述后置滤波（214）步骤中的所述音调后置滤波包括：

为随后的子帧确定由所述随后的子帧的音调值减小的时间索引值；
以及

如果所述确定的值大于或等于当前时间索引，则将所述随后的子帧的所述音调值作为所述当前帧的音调引导参数。

16、根据权利要求 14 或 15 所述的解码方法，包括检测所述基于帧的编码信号的音频特性的步骤；

由此所述后置滤波步骤根据先前帧和后续帧中的至少一个与当前帧之间的音频特性变化的显著性的测量，来调节所述音调参数基于所述后续帧的所述音调参数的程度。

17、根据权利要求 16 所述的解码方法，其中所述检测步骤包括检测语音活动和发声中的至少一个，并且其中所述后置滤波的步骤仅在所检测的浊音语音起始时，将所述音调参数以所述后续帧的所述音调参数为基础。

18、根据权利要求 11 至 17 中任一项所述的解码方法，其中还响应于所述各个后续帧的解码信号执行所述后置滤波（214）步骤。

19、根据权利要求 11 至 18 中任一项所述的解码方法，其中所述解码（212）步骤在可伸缩解码器中进行解码，在该可伸缩解码器中，所述可伸缩解码器的次解码具有比所述可伸缩解码器的主解码高的时延。

20、一种解码方法，其包括至少两种根据权利要求 19 所述的解码方法。

非因果后置滤波器

技术领域

本发明一般涉及音频和/或语音信号的编码和解码，具体地涉及减小编码噪声。

背景技术

通常，音频编码，特别是语音编码执行从模拟输入音频或语音信号到编码域中的数字表示以及返回到模拟输出音频或语音信号的映射，这种数字表示随同表示音频或语音的值或参数的量化或离散化进行。该量化或离散化可被认作是利用编码噪声来扰动真值或参数。音频或语音编码技术将进行这种编码，以使在给定比特率下解码的语音中的编码噪声的影响尽可能小。不过，对语音进行编码所依据的给定比特率限定了编码噪声最好能被减小到的理论极限。目标在于至少使编码噪声尽可能的听不见。

关于编码噪声的适合的观点是假设它是某种附加白噪声或有色噪声。存在一类增强方法，其在解码器处对音频或语音信号进行解码之后，修改编码噪声以使它变得不大听得见，从而导致音频或语音质量得到改进。这种技术通常被称为“后置滤波”，其指的是在实际的解码器后的某后置处理中得到增强的音频或语音信号。存在许多与利用后置滤波器的语音增强有关的公开出版物。一些最基础的文献是〔1〕-〔4〕。

音调后置滤波器的基本工作原理是消除至少部分编码噪声，其充满了浊音语音的谐波之间中的频谱谷。这通常是通过将解码的语音信号与其时移版本进行加权叠加来实现的，其中时移对应于语音的音调延迟或周期。这导致与期望的语音信号，尤其是在语音谐波之间中的语音信号相关的不相关编码噪声的衰减。所描述的效果可以利用非递归式和递归式滤波器结构来获得。在实践中，非递归式滤波器结构是优选的。

本发明的上下文中所关注的是音调或精细结构后置滤波器。它们基本的工作原理是消除充满了浊音语音的谐波之间中的频谱谷的至少部分编码噪声。这通常是通过将解码的语音信号与其时移后版本进行加权叠加来实现的，其中时移对应于语音的音调延迟或周期。优选的是，还

包括成为未来语音信号样本的时移后版本。在〔5〕中描述了一种较近的非递归式音调后置滤波方法，其中在相应的信号样本的后置滤波中再使用信号编码中的音调参数。〔5〕中的非递归音调后置滤波方法也被应用在 3GPP AMR-WB+ 音频和语音编码标准 3GPP TS 26.290, “Audio codec processing functions; Extended Adaptive Multi-Rate-Wideband (AMR-WB+) codec; Transcoding functions (音频编解码器处理功能; 扩展式自适应多速率宽带(AMR-WB+)编解码器; 代码转换功能)”以及 3GPP VMR-WB [3GPP2 C.S0052-A, “Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB), Service Options 62 and 63 for Spread Spectrum Systems (源控制式可变比特率多模式宽带语音编解码器(VMR-WB), 针对扩展频谱系统的服务选项 62 和 63)”中。一种音调后置滤波方法在〔6〕中得到详细说明。该专利描述了对一个并且是同一个帧内的过去和未来的合成语音的使用。

采用评估未来语音信号的后置滤波器的一个问题在于，它们需要访问解码的音频或语音信号的一个未来音调周期。通常可以通过缓冲解码的音频或语音信号来使得该未来信号可用于后置滤波器。不过，由于它增加了编解码器的算法时延，并且因此会影响通信质量尤其是交互性，所以这在音频或语音编解码器的会话式应用中是不期望的。

发明内容

本发明的目的在于提供来自解码器装置的改进的音频或语音质量。本发明的又一目的在于提供与可伸缩解码器装置一起使用的有效的后置滤波装置，该后置滤波装置不会显著促使音频或语音信号的任何附加时延。

以上目的是由根据所附专利权利要求的装置和方法来实现的。一般而言，根据第一方面，一种解码器装置包括用于基于帧的编码信号的参数的接收器输入端，以及连接至该接收器输入端的解码器，该解码器被设置为基于所述参数提供解码音频信号的帧。所述接收器输入端和/或所述解码器被设置为建立在所述接收器输入端可得到第一帧的参数的时刻与在所述解码器的输出端可得到该第一帧的解码音频信号的时刻之间的时间差，该时间差对应于至少一帧。后置滤波器连接至所述解码器的输出端以及所述接收器输入端。所述后置滤波器被设置为响应于各个

后续帧的参数，提供对解码音频信号的帧的滤波以成为输出信号。所述解码器装置还包括用于所述输出信号并连接至所述后置滤波器的输出端。

根据第二方面，一种解码方法包括接收基于帧的编码信号参数以及将所述参数解码成解码音频信号的帧。所述接收和/或所述解码引起在接收后可得到第一帧的参数的时刻与在解码后可得到该第一帧的解码音频信号的时刻之间的时间差，该时间差对应于至少一帧。响应于各个后续帧的参数将所述解码音频信号的帧后置滤波成输出信号。所述方法还包括输出所述输出信号。

本发明的一个优点在于有可能改进语音和音频编解码器的重构信号质量。例如，如果编解码器是可伸缩语音和音频编解码器，或者如果它被用在接收端中具有抖动缓冲器的 VoIP 应用场合中，无需附加时延的任何损失就可获得这种改进。在例如语音起始时的瞬态声期间，特别的增强是可能的。

附图说明

可以通过参照下述结合附图所进行的描述而对本发明和其进一步的目的和优点进行最好的理解，在附图中：

图 1 是具有后置滤波器的音频或语音编解码器的基本结构的示意图；

图 2 示出根据本发明的解码器装置的实施例的框图；

图 3 示出根据本发明的解码器装置的另一实施例的框图；

图 4 是一般的可伸缩音频或语音编解码器的框图；

图 5 是其中较高层支持非语音音频信号的编码的另一可伸缩音频编解码器的框图；

图 6 示出根据本发明的方法的实施例的步骤的流程图；

图 7 示出根据本发明的可伸缩解码器装置的实施例的框图；

图 8 示出根据本发明的可伸缩解码器装置的另一实施例的框图；

图 9 示出根据本发明的可伸缩解码器装置的再一实施例的框图；

图 10 示出根据本发明的可伸缩解码器装置的另一实施例的框图；以

及

图 11 示出根据本发明的改进式音调引导参数 (pitch lead parameter) 计算。

具体实施方式

在本公开的整个内容中，不同附图和实施例中的相同或直接对应的特征将用相同的附图标记来表示。

为了充分理解具体实施方式，一些术语可能必须被限定得更加清楚从而避免混淆。在本公开中，术语“参数”被用作通称，它代表任何类型的信号表示，包括比特或比特流。

为了理解本发明所取得的优势，具体实施方式大体上将从后置滤波的简短综述开始。图 1 示出具有后置滤波器的音频或语音编解码器的基本结构。发送器单元 1 包括将输入的音频或语音信号 3 编码为参数流 4 的编码器 10。参数 4 通常被编码并被传输至接收器单元 2。接收器单元 2 包括解码器 20，该解码器 20 接收表示原始音频或语音信号 3 的参数 4，并将这些参数 4 解码为解码音频或语音信号 5。解码音频或语音信号 5 被规定为与原始音频或语音信号 3 尽可能的相类似。不过，解码音频或语音信号 5 总是包括某种程度上的编码噪声。接收器单元 2 进一步包括后置滤波器 30，该后置滤波器 30 从解码器 20 接收解码音频或语音信号 5，执行后置滤波程序并且输出后置滤波的解码音频或语音信号 6。

后置滤波器的基本思想是对编码噪声的频谱形状进行整形，使得它变得不大听得见，这本质上利用了人类声音感知 (human sound perception) 特性。通常，这会被进行，从而使得噪声在从语音信号具有低功率 (频谱谷) 的区域被移除的同时，被移动到语音信号具有相对高的功率 (频谱峰) 的感知较不敏感的频率区域。存在两种基本的后置滤波器方案，短期和长期后置滤波器，也被称为共振峰和相应的音调或精细结构滤波器。为了获得良好的性能，通常使用自适应后置滤波器。

如上所述，音调或精细结构后置滤波器在本发明中是有用的。解码语音信号与其时移后版本的叠加，导致与期望的语音信号，尤其是在语音谐波之间的语音信号相关的不相关编码噪声衰减。所描述的效果可以利用非递归式和递归式滤波器结构来获得。在 [4] 中描述的一种这样的一般形式由下式给出：

$$H(z) = \frac{1 + \alpha z^{-T}}{1 - \beta z^{-T}}$$

其中 T 对应于语音的音调周期。

在实际中，非递归式滤波器结构是优选的。在公开的美国专利申请 2005/0165603 中描述了一种较近的非递归式音调后置滤波方法，它被应用在 3GPP（第三代合作伙伴计划）AMR-WB+（扩展式自适应多速率宽带编解码器）〔3GPP TS 26.290〕以及 3GPP2 VMR-WB（可变比特率多模式宽带编解码器）〔3GPP2 C.S0052-A：“Source-Controlled Variable-Rate Multimode Wideband Speech Codec (VMR-WB), Service Options 62 and 63 for Spread Spectrum Systems（源控制式可变比特率多模式宽带语音编解码器(VMR-WB)，针对扩展频谱系统的服务选项 62 和 63）”〕音频和语音编码标准中。在这里，基本思想是首先利用下列关系式计算编码噪声估计 $r(n)$ ：

$$r(n) = y(n) - y_p(n)$$

其中 $y(n)$ 是解码音频或语音信号，而 $y_p(n)$ 是如下计算的预测信号：

$$y_p(n) = 0.5 \cdot (y(n-T) + y(n+T)) \quad (1)$$

其次，从语音信号中减去用某个系数 α 加权的噪声估计的低通（或带通）滤波版本，结果产生增强的音频或语音信号：

$$y_{enh}(n) = y(n) - \alpha \cdot LP\{r(n)\} \quad (2)$$

对低通滤波噪声信号的合理解释是，如果在符号上被反向，那么将它看作补偿编码噪声的低频部分的增强信号。响应于预测信号与解码语音信号的相关性、预测信号的能量以及语音信号与预测信号的差分能量的某时间平均值，来修改系数 α 。

如上所述，现有技术中评估以上所定义的表达式 $y_p(n) = 0.5 \cdot (y(n-T) + y(n+T))$ 的音调后置滤波器具有的一个问题在于，它们需要解码语音信号 $y(n+T)$ 的一个未来音调周期，从而依次增加了算法时延。基于可用的解码音频或语音信号并且假定该音频或语音信号将随音调周期 T 周期性地扩展，AMR-WB+ 和 VMR-WB 通过将该解码的音频

或语音信号扩展到未来以解决这个问题。在假定一直到时间索引 $n+$ 且不包括该时间索引 $n+$ 解码的音频或语音信号是可用的情况下，未来音调周期根据以下表达式来计算：

$$\hat{y}(n+T) = \begin{cases} y(n+T) & n+T < n^+ \\ y(n) & n+T \geq n^+ \end{cases}$$

由于这种扩展只是一种近似，因此与在使用真实的未来解码的语音信号时所能获得的相比，在质量上有所折衷。要注意的是，〔6〕也没有提供对于该问题的任何期望的解决方案。而是指出了，只有假定可获得跟随待增强的子帧的子帧时，才在当前帧中利用未来合成语音数据进行后置滤波。特别是，该文献仅想到直到当前语音帧的语音帧的可用性而没有想到未来帧。

另一相关的后置滤波方法被具体描述在〔7〕中，不过，它与本发明上下文具有较低的关联性。该专利描述了一种用于可变速率语音编解码器的后置滤波方法，其中响应于平均比特率来控制该后置滤波的强度。

传统的（例如，共振峰/音调）后置滤波器没有引入任何时延以将编解码器时延保持在最小值。这是因为，编码时延预算通常被更有效地花费在编码器中以用于例如预行估计。这个事实带来下述降低后置滤波器的增强能力的问题。

要注意的是，时间扩展是个问题，尤其是在语音信号的音调周期是非固定的情况下。在浊音语音起始时尤其是这样。更具体地，它可以被这样描述：常规后置滤波器在语音瞬变（speech transient）中的性能由于其参数相比较是不可靠的而不是最优的。

因此，本发明基本思想的一个重要部分是，通过利用来自未来帧的信息增强后置滤波器的性能。为了这样做，采用接收和解码操作中的固有延时。本发明是基于一种这样的情况，其中帧的解码信号变得可用与后续帧的参数变得可用相关或帧的解码信号变得可用晚于后续帧的参数变得可用。换句话说，由接收器输入端和解码器构成的集合（collective）被布置为，提供第一帧 n 的解码信号 $y(n)$ ，基本上同时作为与第一帧 n 相继的帧 $n+1$ 的参数 $x(n+1)$ 。解码的语音帧 $y(n)$ 被送入该后置滤波器中，以产生增强的输出语音帧 $y_{out}(n)$ 。根据本发明，通过向

至少一个后面帧 $n+1$ 的参数 $x(n+1)$ 提供后置滤波器访问来增强后置滤波操作。由于信号时延在接收和解码操作中是固有的，因此并不引起任何附加的信号时延。

一个实施例包括根据算法操作的解码器，该算法引起至少帧长度 L 的输出时延。然后，当解码器输出解码的语音帧 $y(n)$ 时，在接收器中可得到索引 $n+1$ 的编码的语音帧，并且该编码的语音帧能够用于后置滤波的目的。这样的时延在不同的解码器装置中均是存在的。图 2 示出根据本发明的解码器装置的该实施例的框图。接收器单元 2 包括接收器输入端 40，该接收器输入端 40 被设置为接收表示基于帧的编码信号 $x(n+1)$ （通常是编码的语音或音频信号）的参数 4。解码器 20 被连接至接收器输入端 40，其被设置为基于所述参数 4 提供解码音频信号 5 的帧 $y(n)$ 。解码器 20 被设置为表示在接收器输入端 40 可得到第一帧的参数 4 的时刻与在解码器 20 的输出端可得到该第一帧的解码音频信号的时刻之间的时间差，该时间差对应于至少一帧。在本实施例中，解码操作引起一帧的信号时延 51。因此，解码器 20 和接收器输入端 40 的集合 50 同时呈现解码信号 $y(n)$ 作为相继帧 $x(n+1)$ 的参数。

后置滤波器 30 被连接到解码器 20 的输出端以及接收器输入端 40。后置滤波器 30 被设置为响应于后续帧的参数 $x(n+1)$ 而基于解码音频信号的帧 5 提供输出信号 6。从而，可在后置滤波过程中使用未来信号帧的知识，但是却没有增加任何附加解码时延。接收器输出端 60 被连接到后置滤波器 30 以用于输出输出信号 6。

VoIP 系统的一个主要部件是在接收端中的抖动缓冲器。其目的是将所接收到的包含在包中的编码语音帧的异步流转换为随后由语音解码器解码的同步流。因此，抖动缓冲器可根据以上所提出的思想作为参数缓冲器。换句话说，本发明的实施例可以有利地被应用在 VoIP 应用中，其中假定该缓冲器不是空的，接收端中的抖动缓冲器易于提供对未来帧的访问。

因此，本发明另一实施例包括接收器输入端，该接收器输入端又包括参数缓冲器，该参数缓冲器存储所接收到的编码语音帧，至少两帧。解码器对缓冲的帧 n 进行解码，从而产生解码语音帧 $y(n)$ 。同时，索引 $n+1$ 的编码语音帧在参数缓冲器中是可得到的，并可以用于后置滤波的目的。图 3 示出根据本发明的解码器装置的这种实施例的框图。接收器

单元 2 包括接收器输入端 40，该接收器输入端 40 被布置为接收表示基于帧的编码信号参数 4。接收器输入端 40 包括抖动缓冲器 41，该抖动缓冲器 41 具有用于至少两帧的参数的存储位置 42A、42B。

解码器 20 被连接到抖动缓冲器 41 的第一位置 42A，从而被提供有第一帧 $x(n)$ 的参数 4A。解码器 20 被设置为基于参数 4A 提供解码音频信号 5 的帧 $y(n)$ 。接收器输入端 40 由于抖动缓冲器 41 而呈现下述时间差，该时间差是在接收器输入端 40 可得到特定帧的参数 4B 的时刻与在解码器 20 的输出端可得到相同帧的解码音频信号 5 的时刻之间的时间差，该时间差对应于至少一帧。在本实施例中，抖动操作引起至少一帧的信号时延。因此，解码器 20 和接收器输入端 40 的集合 50 同时呈现解码信号 $y(n)$ 作为相继帧 $x(n+1)$ 的参数。然后，以与图 2 中相同的方式设置后置滤波器 30。

图 4 示出根据本发明的方法的实施例的步骤的流程图。该解码方法在步骤 200 开始。在步骤 210，接收基于帧的编码信号参数。在步骤 212，将参数解码为解码音频信号的帧。步骤 210 和 212 中至少之一引起在接收后可得到第一帧的参数时刻与在解码后可得到该第一帧的解码音频信号时刻之间的时间差。该时间差对应于至少一帧。在步骤 214，响应于各个后续帧的参数，将解码音频信号的帧后置滤波为输出信号。在步骤 216，输出该输出信号。该流程在步骤 299 结束。

具有固有时延的编解码器的典型示例是可伸缩或嵌入式编解码器。因此，下文介绍可伸缩编解码器的简短评述。图 5 示出一般的可伸缩音频或语音编解码器系统的框图。此处的发送器单元 1 包括编码器 10，在这种情况下编码器 10 是将输入的音频或语音信号 3 编码为参数流 4 的可伸缩编码器 110。整个编码发生在两个层中，即在包含主编码器 11 的发送器中的下层 7，以及在包含次编码器 15 的发送器单元中的至少一个上层 8。可伸缩编解码器装置可具有附加层，但是在本公开中使用两层解码器系统作为模型系统。不过，本发明的原理还可以被应用于具有两个以上层的可伸缩编解码器。主编码器 11 接收输入的音频或语音信号 3，并将它编码为主参数 12 流。该主编码器还将主参数 12 解码为估计主信号 13，该信号理论上会对应于可在解码器侧从主参数 12 获得的信号。在比较器 14（在此情况下是减法单元）中比较估计主信号 13 与原始输入的音频或语音信号 3。因此，差信号是主编码器 11 的主编码噪声

信号 16。主编码噪声信号 16 被提供至次编码器，该次编码器将该信号 16 编码为次参数 17 流。这些次参数 17 可被看作可从主参数 12 中解码的优选增强信号的参数。主参数 12 和次参数 17 一起形成输入音频或语音信号 3 的总参数 4 流。

参数 4 通常被编码并被传输至接收器单元 2。接收器单元 2 包括解码器 20，在这种情况下是可伸缩解码器 120，该解码器 120 接收表示原始音频或语音信号 3 的参数 4，并将这些参数 4 解码为解码音频或语音信号 5。整个解码也发生在两个层中，即下层 7 和上层 8 中。在该接收器单元中，下层 7 包括主解码器 21。类似地，上层 8 包括在接收器单元中的次解码器 25。主解码器 21 接收参数 4 流的输入主参数 22。理论上，这些参数与编码器 10 中建立的那些参数相同，不过，在有些情况下，传输噪声可能使参数失真。主解码器 21 将输入的主参数 22 解码为解码的主音频或语音信号 23。类似地，次解码器 25 接收参数 4 流的输入次参数 27。理论上，这些参数与编码器 10 中建立的那些参数相同，不过，在有些情况下，此处的传输噪声也可能使参数失真。次解码器 21 将输入的次参数 22 解码为解码的增强音频或语音信号 26。该解码的增强音频或语音信号 26 被规定为与主编码器 11 的编码噪声尽可能准确地对应，从而也与由主解码器 21 产生的编码噪声相似解码的主音频或语音信号 23 和解码的增强音频或语音信号 26 在加法器 24 中相加，从而得到最后的输出信号 5。

如果在接收单元 2 中仅接收到主参数 22，那么该接收单元仅支持主解码，或者次解码无论如何都被确定为不执行，所得到的解码的增强音频或语音信号 26 将等于零，而且输出信号 5 将变成与解码的主音频或语音信号 23 相同。这显示出可伸缩编解码器系统的概念的灵活性。任何后置滤波通常都可根据现有技术输出信号 5 上执行。

现如今使用最多的可伸缩语音压缩算法是根据 ITU-T 建议 G.711，1988 年 11 月，“Pulse code modulation (PCM) of voice frequencies on a 64kbps channel (64kbps 信道上的声音频率的脉冲编码调制 (PCM))”的 64kbps 的 A/U-律对数 PCM 编解码器。8kHz 采样的 G.711 编解码器将 12 比特或 13 比特的线性 PCM (脉冲编码调制) 样本转换成 8 比特对数样本。对数样本的有序比特表示允许窃取 G.711 比特流中的最低有效位 (LSB)，使 G.711 编码器实际上在 48、56 和 64kbps 之间是 SNR (信

噪比)可伸缩的。G.711 编解码器的这种可伸缩特性被用在电路交换通信网络中以用于带内控制信令的目的。使用该 G.711 伸缩特性的近期示例是 3GPP-TFO 协议 (TFO 为根据 3GPP TS28.062 的无汇接操作), 其现在保留 64kbps PCM 链路上的宽带语音设置和传输。最初使用原始 64kbps G.711 流中的 8kbps 以允许宽带语音服务的呼叫设置, 而不会对窄带服务质量有很大影响。在呼叫设置之后, 宽带语音将使用 64kbps G.711 流中的 16kbps。其它支持开环可伸缩性的老式语音编码标准是 ITU-T 建议 G.727, 1990 年 12 月, “5-, 4-, 3- and 2-bit/sample embedded adaptive differential pulse code modulation (ADPCM) (5-、4-、3-和 2-比特/样本嵌入式自适应差分脉冲编码调制 (ADPCM))”以及在某种程度上的 G.722 (子带 ADPCM)。

可伸缩语音编码技术中的较近发展是为 MPEG4-CELP 提供可伸缩性扩展的 MPEG-4 (运动图像专家组) 标准 (ISO/IEC-14496)。通过附加滤波器参数信息或附加新息 (innovation) 参数信息的传输, MPE 基层可以得到增强。国际电信联盟标准化部门, ITU-T 最近根据 ITU-T 建议 G.729.1, 2006 年 5 月, “G.729 based Embedded Variable bit-rate coder: An 8-32 kbit/s scalable wideband coder bitstream interoperable with G.729 (基于 G.729 的嵌入式可变比特率编码器: 与 G.729 共同操作的 8-32 千比特/秒可伸缩宽带编码器比特流)”, 简称为 G.729.EV, 已经结束了新的可伸缩编解码器的标准化。该可伸缩语音编解码器的比特率范围为从 8kbps 到 32 kbps。该编解码器提供 8-32kbps 的可伸缩性。该编解码器的主要用途是允许在家庭或办公网关中有效共享有限带宽资源, 例如, 在若干 VoIP 电话 (互联网协议电话) 之间的共享 xDSL 64/128 kbps 上行链路 (DSL 为数字用户线路, xDSL 为各种特定 DSL 方法的通称)。

可伸缩语音编码的一个最近趋势是提供支持例如音乐的非语音音频信号的编码的较高层。图 6 中示出一种这样的方法。在这种编解码器中, 低层 7 采用纯常规的语音编码, 例如根据 CELP (码激励线性预测) 是突出示例的综合分析 (AbS) 范例。因此, 在本实施例中, 主编码器 11 是 CELP 编码器 18, 而主解码器 21 是 CELP 解码器 28。由于这种编码仅仅对于语音是特别适合的, 而对于例如音乐的非语音音频信号就不是那么适合, 因此上层 8 实际上根据音频编解码器中使用的编码范例来操作。因此, 在本实施例中, 次编码器是音频编码器 19, 而次解码器是

音频解码器 29。在本实施例中，上层 8 编码通常对下层编码的编码误差起作用。

图 7 中所示的本发明的一个特定实施例是在可伸缩语音/音频解码器 120 中的应用，在该可伸缩语音/音频解码器 120 中，低层在主解码器 21 中执行主解码得到主解码信号 y_p ，而较高层在次解码器 25 中执行次解码得到次增强信号 y_s 。次增强信号 y_s 将主解码信号 y_p 改进为增强的解码信号 y_e 。在本实施例中，假定解码器 20 在例如 20 毫秒长度的语音帧上操作，并且主解码器 21 具有比次解码器 25 低至少一帧的时延。换句话说，固有时延 51 存在于次解码器 25 中。

在一些特殊的编解码器系统中，次编解码器可以利用与主编解码器不同的帧长度操作。例如，次编解码器可以具有较之主编解码器一半的帧长度，因此它的主解码器解码一帧的同时可解码两个次帧。依据于设计，次解码器的固有时延或者是主解码器的帧长度，或者是次解码器的帧长度。

具体地且如在图 7 中所见的，假定主解码器 21 可以将第 $n+1$ 语音帧 $x(n+1)$ 解码为主解码信号 23 的输出帧 $y_p(n+1)$ 而没有任何特定的时延，也就是，基于相应的具有帧索引 $n+1$ 的所接收到的编码语音帧数据 $x(n+1)$ 。相反地，次解码器 25 甚至需要下一解码帧数据。因此，采用具有索引 $n+1$ 的可用帧 $x(n+1)$ ，次解码器 25 输出解码的次增强信号 26 的解码帧 $y_s(n)$ 。为了适当地组合解码的次增强信号 26 与主解码信号 23，后者必须被延迟一帧。这在时延滤波器 53 中执行，并得出时延的解码主信号 54。

该事实使得可以应用本发明，而不会有甚至进一步增加解码器中的时延的任何损失，这种损失可能是所不期望的。如果所接收到的比特流包含增强层信息，那么可以产生解码的次增强信号 26 的帧 $y_s(n)$ 。该信号 26 与时延的主解码信号的帧 $y_p(n)$ 相组合，从而一起形成增强的解码信号的帧 $y_e(n)$ 。当参数的帧 $x(n+1)$ 变得可从集合 50B 得到时，该帧 $y_e(n)$ 变得可用。帧 $y_e(n)$ 随后可被馈给通过非因果次后置滤波器 30B，其可以利用本发明的优点，如上面进一步所描述的那样。可根据这些思想通过利用帧 $n+1$ 的编码参数而改进后置滤波器 30B 的操作。此外，该后置滤波器 30B 可进一步从利用主解码信号 23 的下一帧 $y_p(n+1)$ 中得到益处，该下一帧构成了仍不能获得的未来帧 $y_e(n+1)$ 的近似。因此，在本实施例中，

后置滤波器 30B 不仅能够基于未来帧的参数来增强信号,而且还能够根据未来帧的实际信号的相当好的近似来增强信号。这样,次后置滤波器 30B 提供后置滤波的增强信号 56 作为来自解码器装置的输出信号 6。

图 8 示出根据本发明的可伸缩解码器装置的另一实施例的框图。在该实施例中,主后置滤波器 30A 被提供,并被连接至时延滤波器 53 的输出端,即其对时延的解码主信号 54 进行操作。在该实施例中,集合 50A 包括接收器输入端 40、主解码器 21 和时延滤波器 53。主后置滤波器 30A 根据本发明进行操作以有权访问后来帧的参数。在该实施例中,相继帧的解码主信号 23 也是可用的,并且还能有利地被用在主后置滤波器 30A 中。换句话说,时延的解码主信号 54 的语音帧 $y_p(n)$ 可通过非因果主后置滤波器 30A 得到增强,这利用了它对解码主信号 23 的语音帧 $y_p(n+1)$ 以及对帧 $n+1$ 的参数 4 的访问。

后置滤波器 30A 的输出信号 55, 即 $y_p^*(n)$ 被用于与次增强信号 26 相组合以产生最后的输出信号。不过,在一些状况下,次增强信号 26 所提供的增强可能在某些情况下与主后置滤波器 30A 所能获得的相似,并且结果可能是编码噪声的过补偿。后置滤波器 30A 在这种情况下可以有利地被设置为用于确定次解码的参数在接收器输入端 40 处是否可用。如果次参数是可用的,那么后置滤波器的操作可以被切断,从而给出原始解码主信号作为主后置滤波器 30A 的输出,或者至少改变后置滤波原理,以便不干扰次增强信号的操作。

图 9 示出根据本发明的可伸缩解码器装置的又一实施例的框图。在该实施例中,次解码器 25 又如图 7 中所示的那样后面跟随着次后置滤波器 30B, 不过,也提供了主后置滤波器 30A。在这样的实施例中,被提供具有来自次解码器 25 的增强的输出信号也可以被利用次后置滤波器 30B 进行进一步增强。而且在这种情况下,次后置滤波器 30B 可将其操作基于相继帧的参数。虽然该后置滤波器 30B 无法访问增强的解码器输出 5 的未来帧 $y_e(n+1)$, 但它的操作实际上可以基于主解码信号的未来帧 $y_p(n+1)$ 。主集合 50A 包括接收器输入端 40、主解码器 21 和时延滤波器 53, 而次集合 50B 包括接收器输入端 40、整个可伸缩解码器 120 和主后置滤波器 30A。

图 10 示出根据本发明的可伸缩解码器装置的又一进一步实施例的框图。此处,非后置滤波时延的解码主信号 54 被提供至加法器 24, 以

与次增强信号 26 相组合。这避免了将主后置滤波器 30A 的编码噪声校正与来自次解码器 25 的增强混合。相反,输出端 60 被设置为选择器 61,该选择器被设置为输出后置滤波的解码主信号 55 或者后置滤波的增强信号 56 作为来自解码器装置的输出信号。选择器 61 优选地响应于输入的信号进行操作,如虚线箭头 62 所示的那样。更多的这些可能性在下面被进一步论述。

如上文所述,本发明的进一步部分方面将依据语音或音频信号的特性来应用后置滤波器的非因果增强。尤其是,这种应用在声音瞬变期间是有益的。这种声音瞬变例如为从一个电话(语音部件)到另一个电话的转变,这些电话本身是相对稳定或固定的。这种瞬变典型的为信号是非固定的,并且由语音编码器所完成的参数估计不如在稳定声音期间可靠。如果后置滤波器基于这种较不可靠的参数,那么有可能的是它的性能是差的。根据本发明,在这种瞬变期间的后置滤波器性能可通过利用未来帧的参数以及优选地还有未来帧的合成语音而得到改进。由于在未来帧期间的声音可以变得更稳定从而允许更可靠的参数估计,因此这种改进可得以实现。

该实施例依赖于瞬变的检测,其中启用了特定的非因果后置滤波操作。这种检测可以利用声音分类器进行,在简单情况下,该声音分类器可以是语音活动检测器(VAD),或者更具体的是声音检测器,该声音检测器除了基本的语音/非语音辨别之外,可以例如区别类似浊音、清音的不同类型的语音的起始。这种检测也可以基于对例如能量或 LPC 参数的特定信号参数的时间演化的评估,并且将语音或音频信号的该部分识别为这些参数迅速变化的瞬变。瞬变检测器可以在编码器或解码器中实现,在前一种情况下需要向接收器传送检测信息。音频特性的变化可以被量化为显著性程度(significance degree)且可以被测量,并用于控制后置滤波器的操作。特别是,根据本发明的后置滤波器可以被设置为调节音调后置滤波器中所使用的音调参数基于后续帧的音调参数的程度。根据当前帧与先前帧或后续帧之间的音频特性变化的显著性的测量来执行该调节。

后置滤波器性能可得到改进的一个特定优选实施例是在语音休止周期之后对浊语音起始的应用。在这里,特别地,后置滤波器是音调后置滤波器,并且它所使用的来自未来帧的参数是属于跟随当前帧的那一

帧的子帧音调参数。

根据进行音调后置滤波器改进的本发明的一个进一步优选的实施例，以新颖且更精确的方式来处理音调参数。如上所述，技术现状是音调后置滤波器基于等式(1)和(2)评估表达式，其中合成语音的过去和未来片段与当前的语音片段相组合，其中片段可以是如子帧或音调周期之类的单元。这些过去和未来的片段延迟分别引导具有音调参数值 T 的当前片段。使用 T 作为过去语音片段的延迟参数在概念上是正确的，这是因为它符合典型的综合分析语音编解码器的自适应代码本搜索范例，该综合分析语音编解码器计算 T 作为延迟值，该延迟值使延迟后片段与当前语音片段的相关性最大化。

不过，由于假定音调延迟参数即使对于未来片段也保持恒定，因此使用 T 作为未来片段的引导参数通常是不精确的。这在音调可能剧烈变化的瞬变中尤其是有问题的。参考文献[6]通过基于片段之间的相关性计算来指定附加延迟和引导决定因素，提供了对这种问题的解决方案。不过，由于复杂性的理由，这是不利的。

根据本发明的对这种问题的解决方案如下参照图 11 所述。假定音调后置滤波器有权访问针对当前帧 n 和至少一个未来帧 $n+1$ 的子帧音调参数的向量。通常，每个帧包含 4 个子帧。 $T[0]..T[3]$ 表示当前帧的四个子帧音调参数，而 $T[4]..T[7]$ 表示未来帧的四个子帧音调参数。假定，通过搜索该子帧音调参数来找到给定片段的引导参数，该子帧音调参数与它在时间延迟中所处的当前片段中的子帧位置相关。根据图 11 中对于给定的当前片段 100 的示例，子帧音调值 $T[4]$ 就是这样的情况。如该图中还可以看到，由于音调正变化成更小的值，因此使用当前片段 $T[1]$ 的音调参数值作为引导参数是不精确的。如下参照图 12 的是优选的示例算法，根据该算法可以找到给定片段的引导参数。将是图 4 中的步骤 214 的一部分的流程在步骤 220 开始。在步骤 222，选择跟随当前片段的子帧。从跟随当前片段的该子帧开始，在步骤 224，检验由相应的子帧音调值减少的子帧时间索引是否大于或等于当前片段的时间索引。如果是这样，则在步骤 226 将子帧音调值当作当前片段的音调引导参数，并且在步骤 239，该算法停止。否则的话，对下一子帧重复该检验。在步骤 228，检验是否存在更多可用的子帧。如果没有，则在步骤 239 该流程结束，否则的话，在步骤 230 选择新的子帧，并且重复

步骤 224 的检验。在该算法中，子帧时间索引可以是，例如子帧的起始或中间时间索引。需要注意的是，如果参考文献 [6] 中所描述的引导决定因素被使用，那么也能使用具有某增益的这种算法，因为这样能够通过限制必须被执行的相关计算所在的范围来帮助节省复杂度。

以上所描述的实施例应该被理解为本发明的一些示例性示例。本领域技术人员将理解的是，可以对实施例进行各种修改、组合和改变，而不偏离本发明的范围。特别是，不同实施例中的不同部分方案可以在技术上可能存在的其它结构中被结合。不过，本发明的范围是由所述权利要求书来限定的。

参考文献

[1] P. Kroon, B. Atal, "Quantization procedures for 4.8 kbps CELP coders", in Proc IEEE ICASSP, pp.1650-1654, 1987 (P. Kroon, B. Atal, "4.8kbps CELP 编码器的量化程序", IEEE ICASSP 会刊, 第 1650-1654 页, 1987 年)。

[2] V. Ramamoorthy, N.S. Jayant, "Enhancement of ADPCM speech by adaptive postfiltering", AT&T Bell Labs Tech. J., pp. 1465-1475, 1984 (V. Ramamoorthy, N.S. Jayant, "利用自适应后置滤波的 ADPCM 语音的增强", AT&T 贝尔实验室技术期刊, 第 1465-1475 页, 1984 年)。

[3] V. Ramamoorthy, N.S. Jayant, R. Cox, M. Sondhi, "Enhancement of ADPCM speech coding with backward-adaptive algorithms for postfiltering and noise feed-back", IEEE J. on Selected Areas in Communications, vol.30 SAC-6, pp. 364-382, 1988 (V. Ramamoorthy, N.S. Jayant, R. Cox, M. Sondhi, "利用用于后置滤波的向后自适应算法和噪声反馈的 ADPCM 语音编码的增强", 所选通信区域 IEEE 期刊, 1988 年第 30 卷 SAC-6, 第 364 - 382 页)。

[4] J. H. Chen, A. Gersho, "Adaptive postfiltering for quality enhancements of coded speech", IEEE Trans. Speech Audio Process., vol. 3, no. 1, 1995 (J. H. Chen, A. Gersho, "用于编码后语音的质量增强的自适应后置滤波", IEEE 语音音频处理学报, 1995 年第 1 期第 3 卷)。

[5] B. Besette et al., "Method and device for frequency-selective pitch enhancement of synthesized speech", Patent application US20050165603A1

(B. Besette 等, “用于合成语音的选择频率音调增强的方法和装置”, 专利申请 US20050165603A1) .

[6] L. Bialik et al., "A pitch post-filter", EP-0807307B 1 (L. Bialik 等, “音调后置滤波器”, EP-0807307B 1) .

[7] Pasi Ojala et al., "A decoding method and system comprising an adaptive postfilter", EP 1 050 040 B1 (Pasi Ojala 等, “包括自适应后置滤波器的解码方法和系统”, EP 1 050 040 B1) .

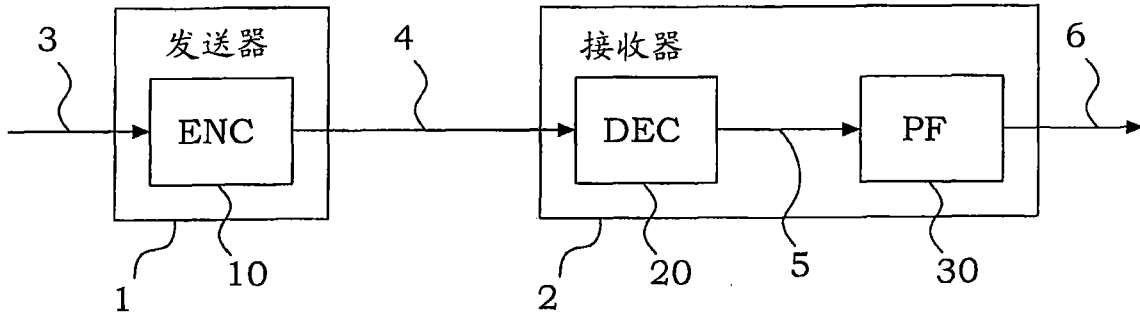


图 1

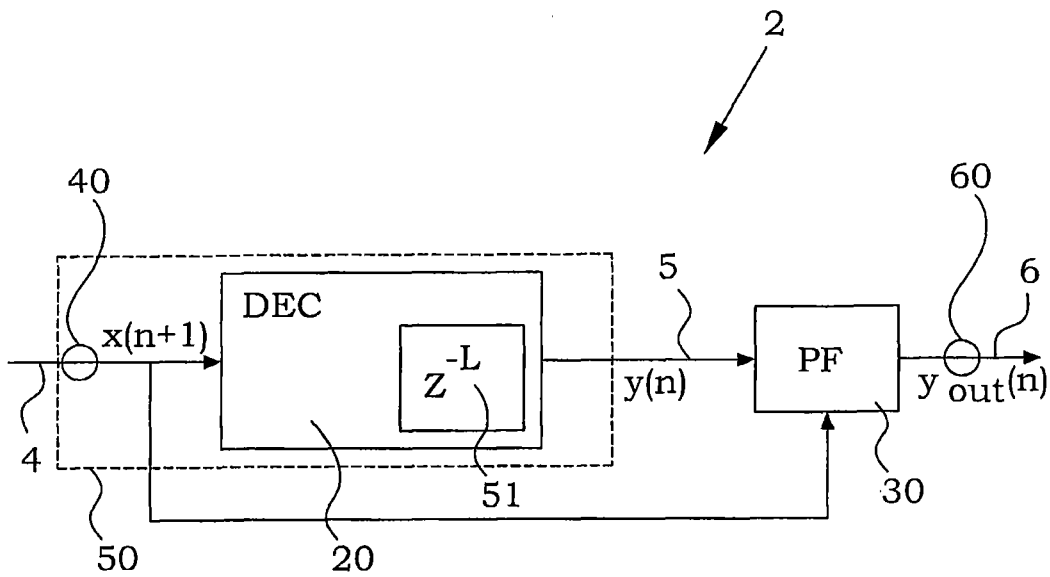


图 2

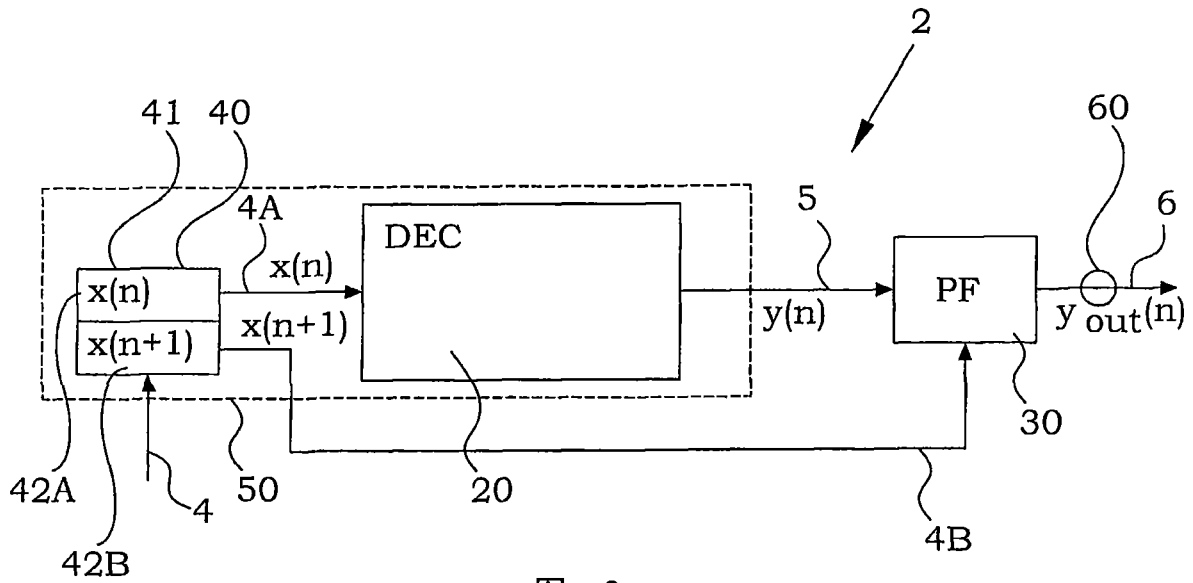


图 3

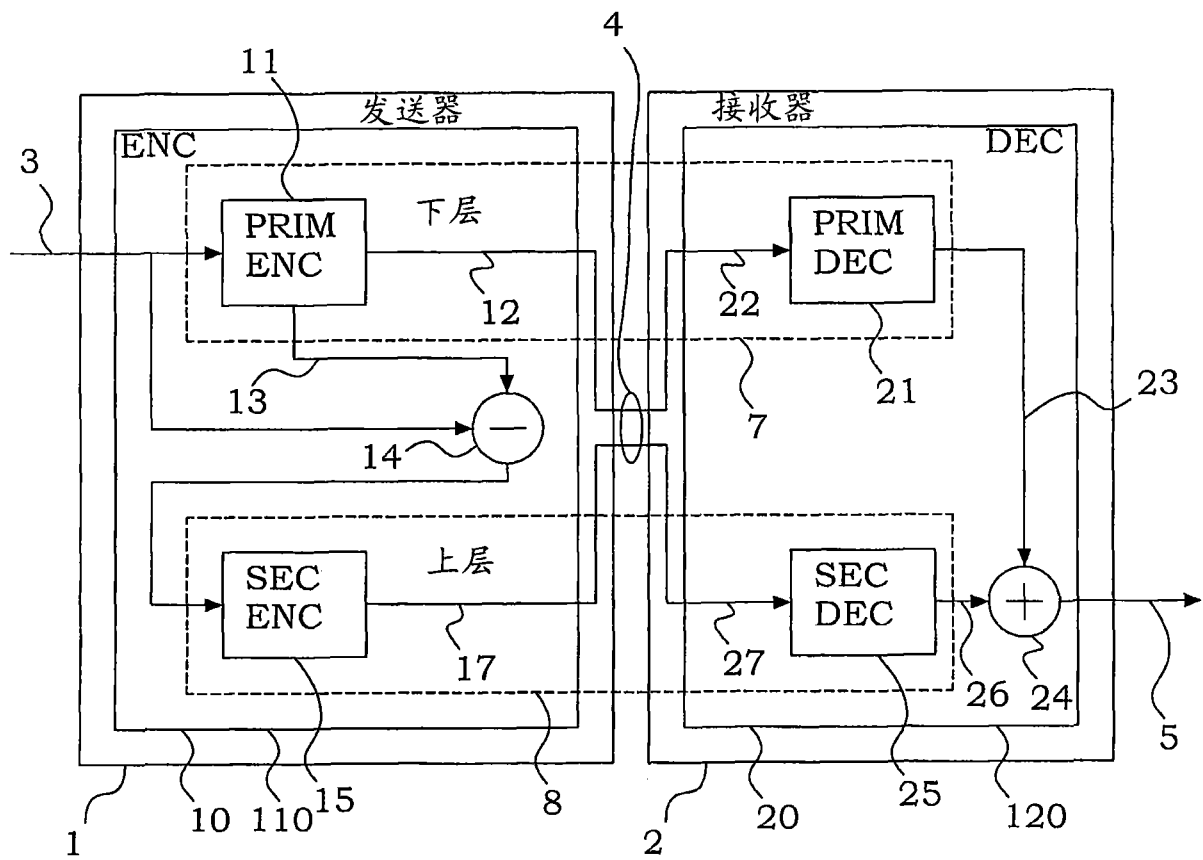


图 5

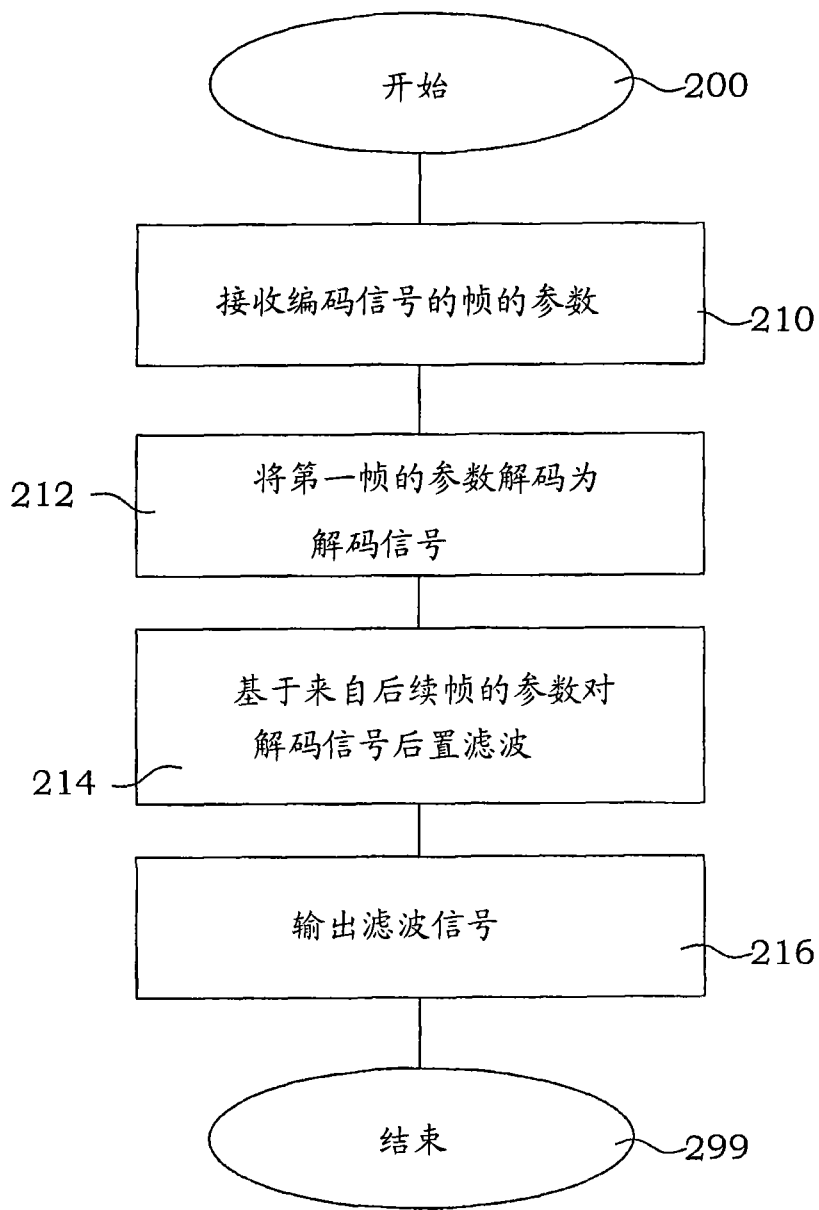


图 4

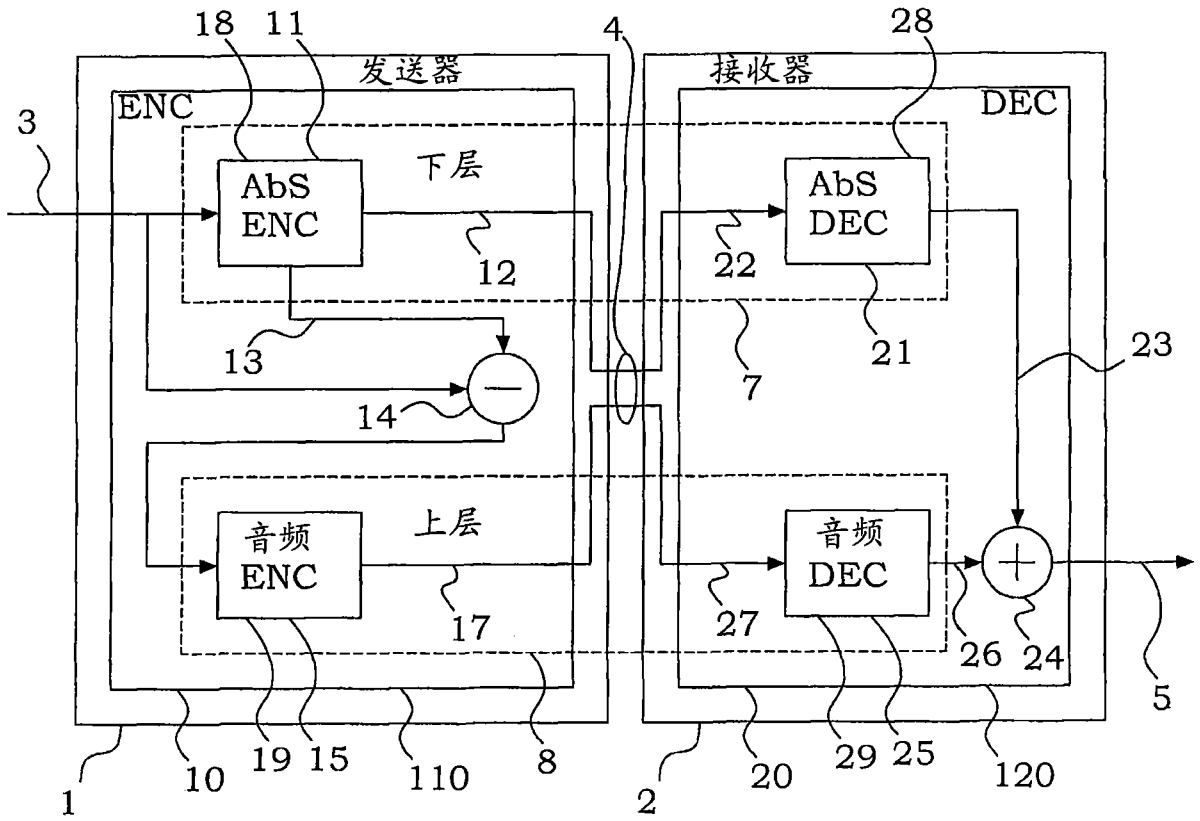


图 6

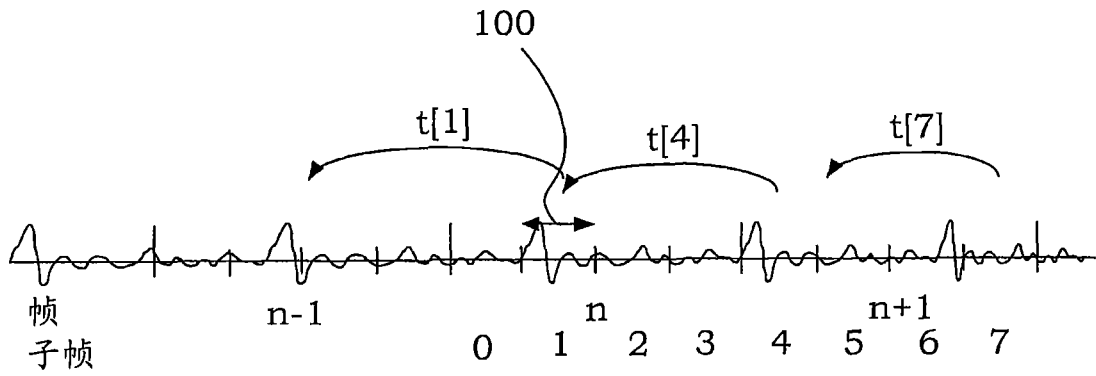


图 11

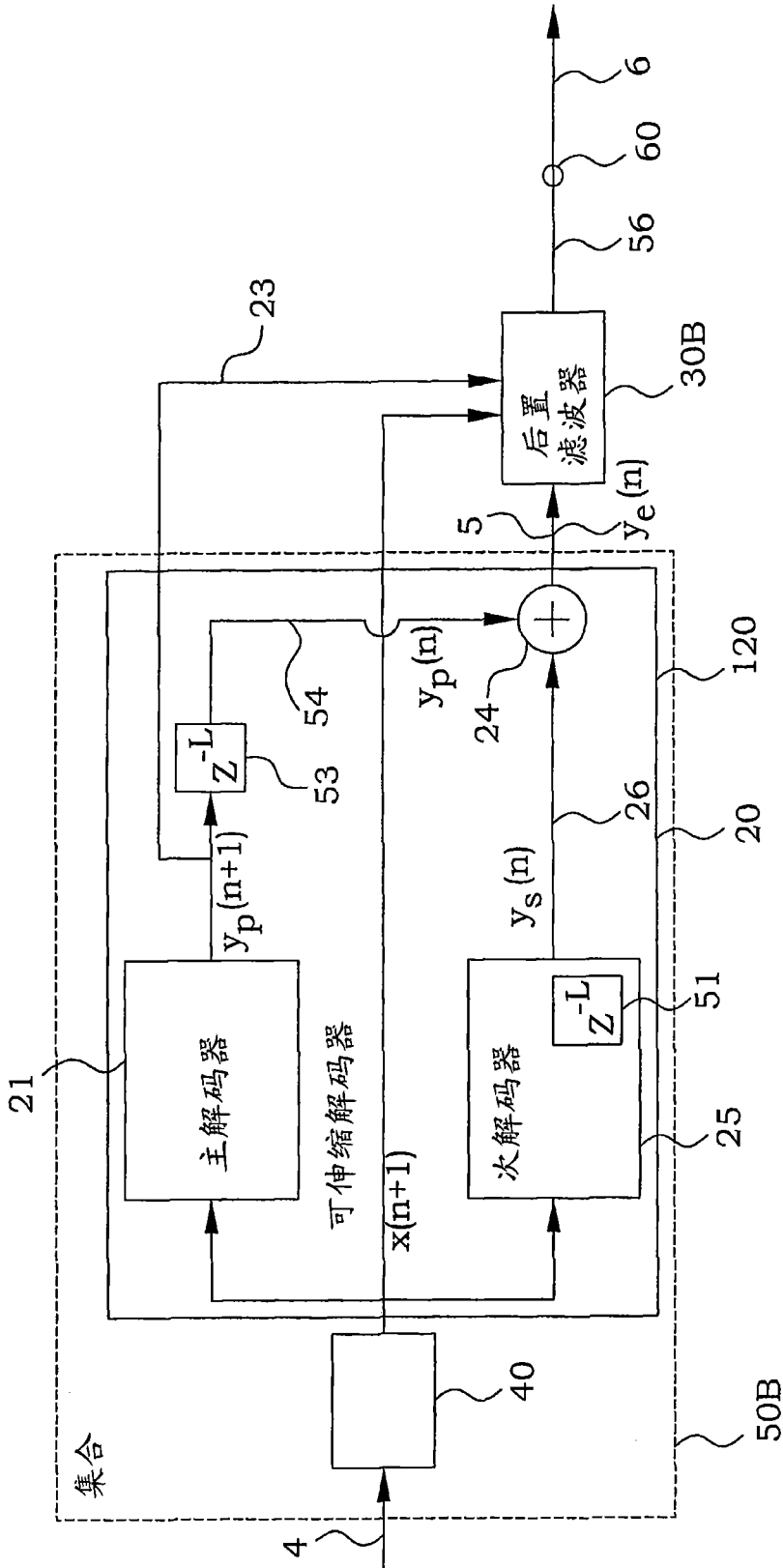


图 7

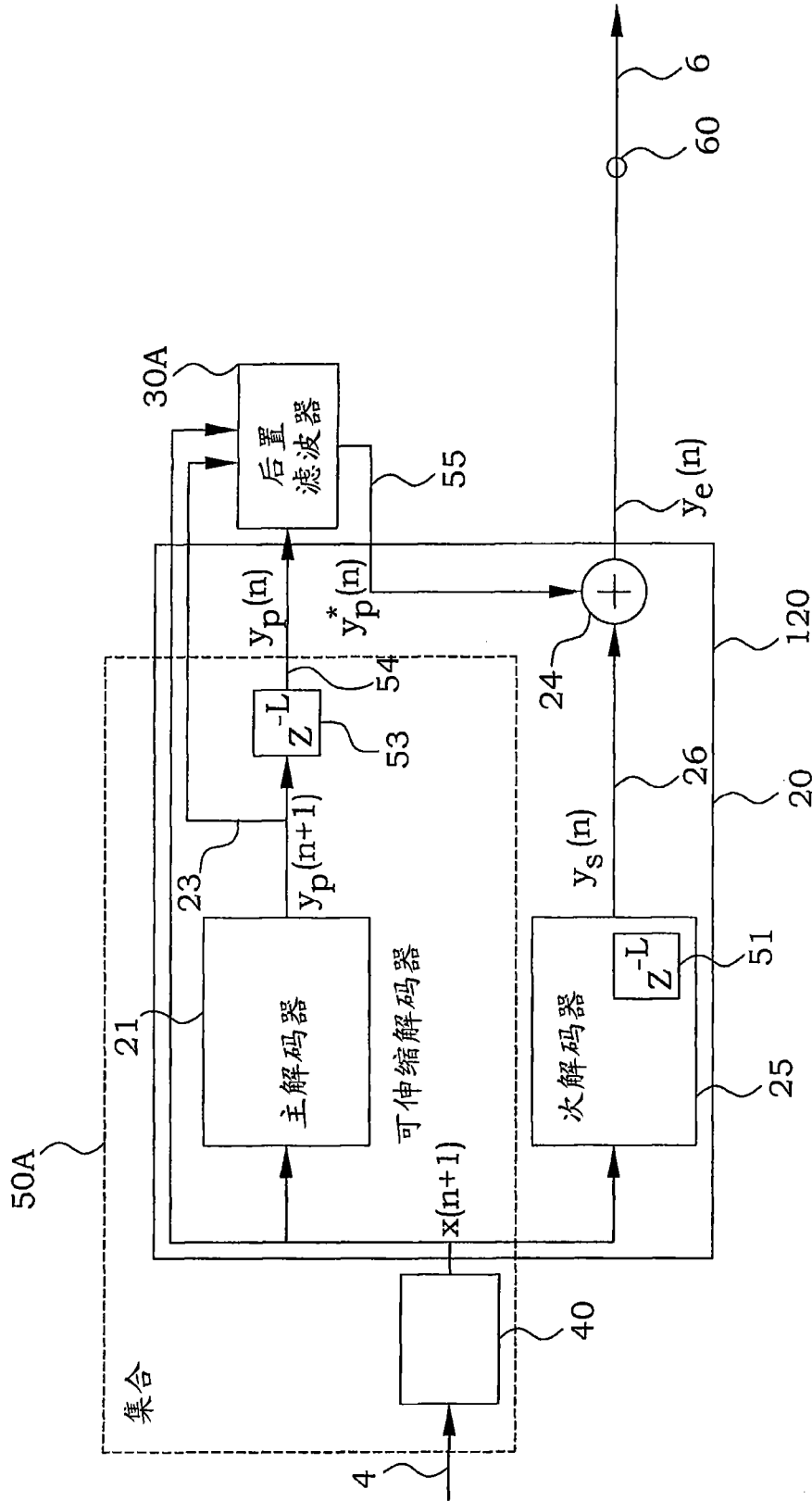


图 8

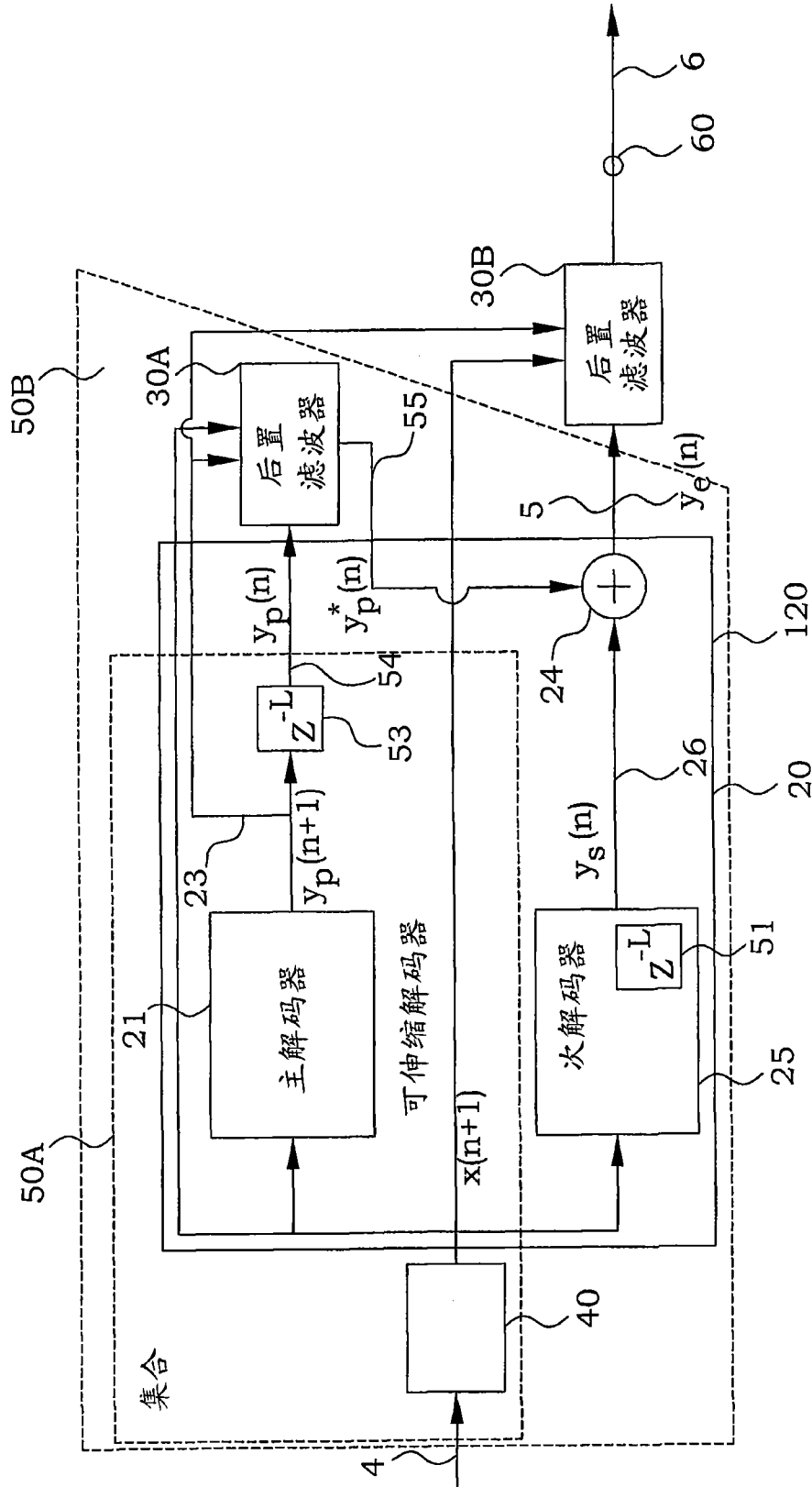


图 9

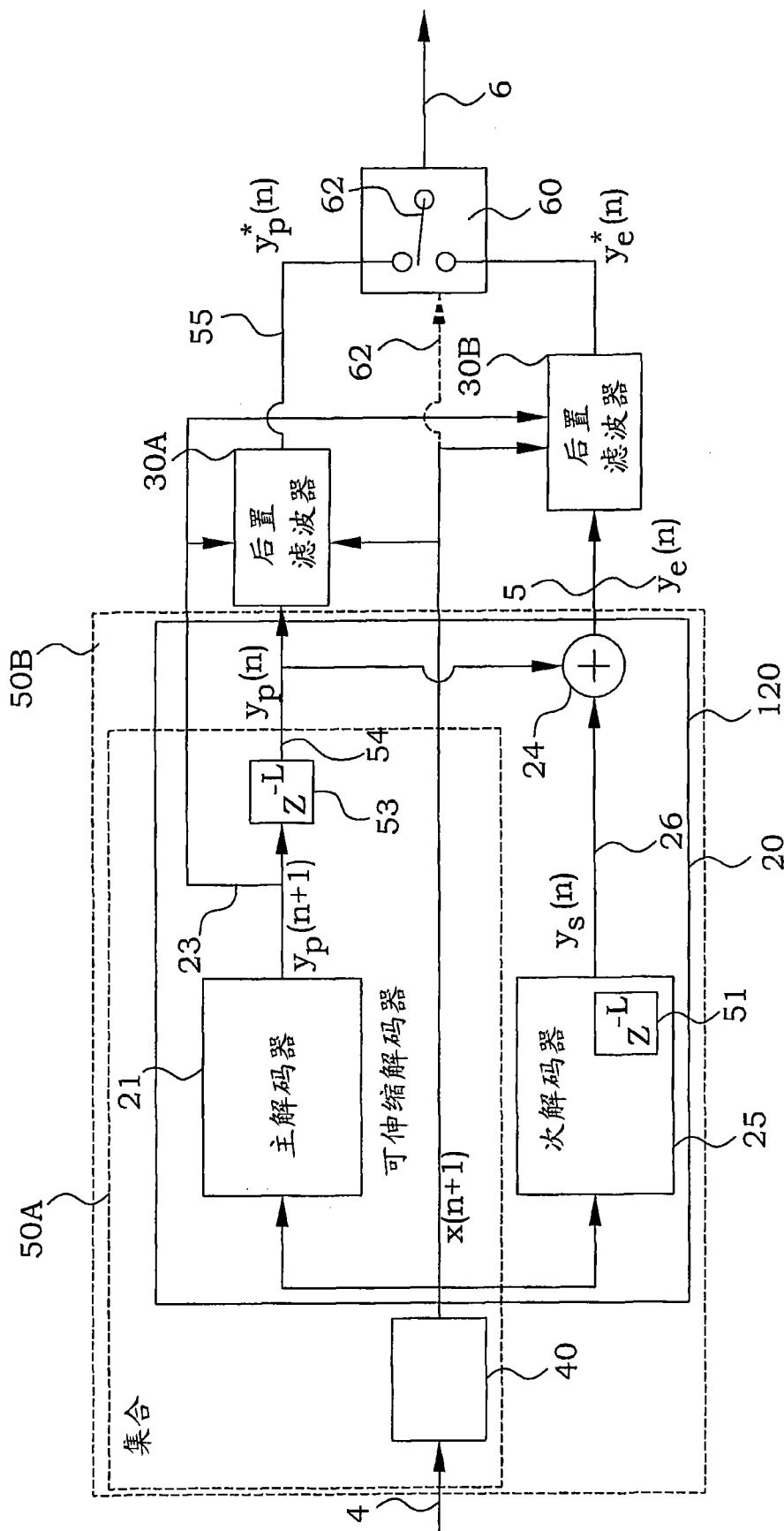


图 10

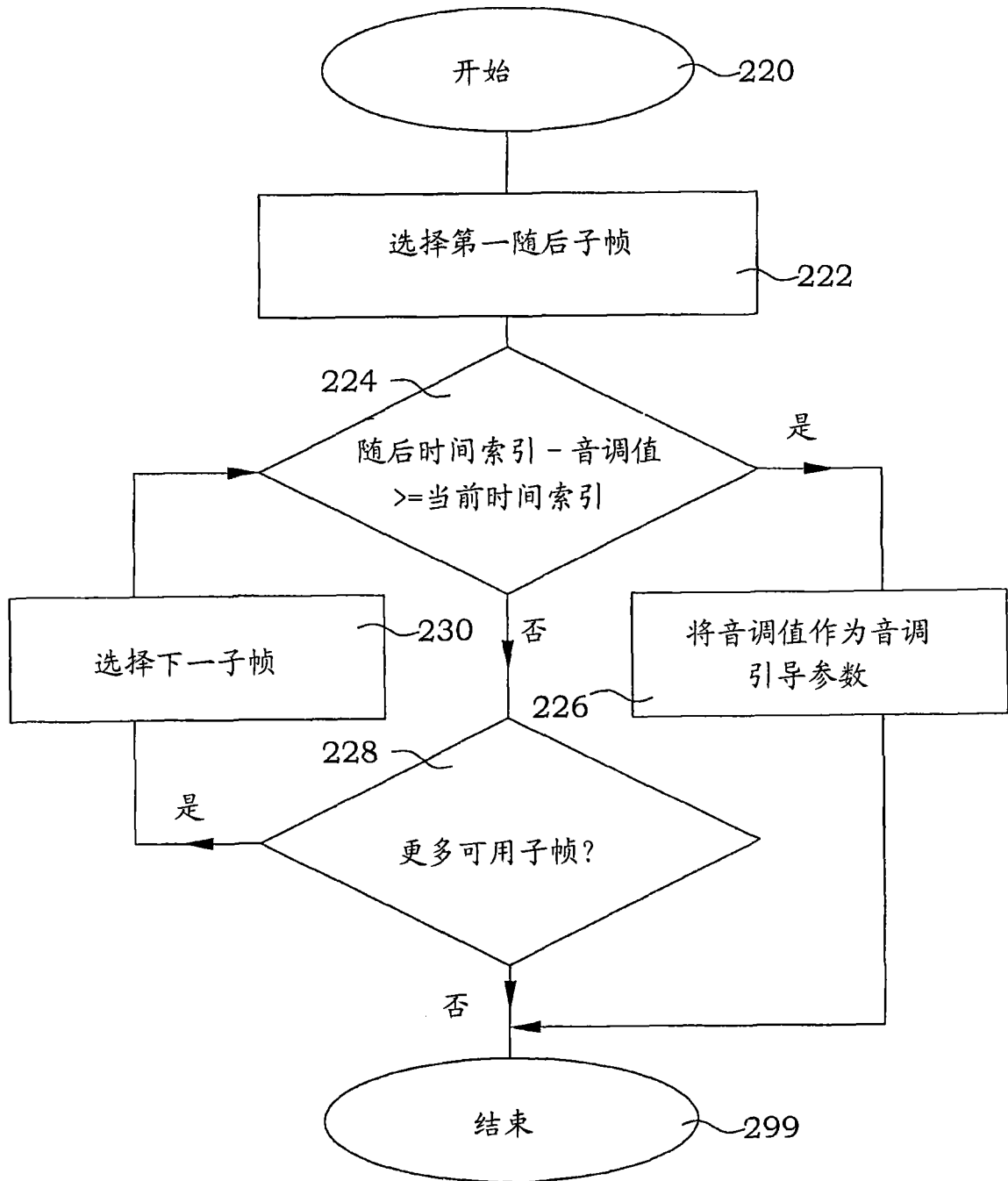


图 12