



## (12) 发明专利

(10) 授权公告号 CN 1975859 B

(45) 授权公告日 2012.06.20

(21) 申请号 200610143655.X

(22) 申请日 2006.10.27

(30) 优先权数据

0522051.2 2005.10.28 GB

(73) 专利权人 索尼欧洲有限公司

地址 英国萨里郡

(72) 发明人 W·E·C·肯蒂什 N·J·海恩斯

(74) 专利代理机构 北京东方亿思知识产权代理  
有限责任公司 11258

代理人 宋鹤

(51) Int. Cl.

G10L 19/00 (2006.01)

G01R 29/26 (2006.01)

H03K 3/84 (2006.01)

G06K 9/00 (2006.01)

H04N 5/913 (2006.01)

(56) 对比文件

CN 1643592 A, 2005.07.20, 全文.

US 6061793 A, 2000.05.09, 全文.

US 2005/0137876 A1, 2005.06.23, 全文.

US 2005/0043830 A1, 2005.02.24, 全文.

Benito Carnero et.al. Perceptual  
Speech Coding and Enhancement Using  
Frame-Synchronized Fast Wavelet Packet  
Transform Algorithms.《IEEE TRANSACTIONS ON  
SIGNAL PROCESSING》.1999, 第47卷(第6期),  
全文.

审查员 辛杰

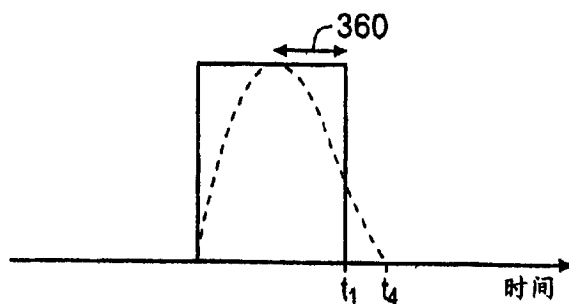
权利要求书 1 页 说明书 12 页 附图 15 页

(54) 发明名称

音频处理方法和装置

(57) 摘要

一种将有效载荷信号插入到原始音频信号中的音频处理设备,它包括:噪声发生器,可用于根据有效载荷信号生成噪声信号;电平检测器,用于检测原始信号的信号电平;调制器,用于响应原始音频信号中检测的信号电平的增加或降低而增加或降低噪声信号的电平,以便生成调制的噪声信号;组合器,用于将原始信号与调制的噪声信号组合;以及信号延迟设置;该调制器相对于信号延迟设置来操作,使得噪声信号的电平中的降低相对于原始音频信号的信号电平中的对应降低是时间超前的。



1. 一种音频处理设备,其中将基于有效载荷信号的信号插入到原始音频信号中,所述设备包括:

噪声发生器,可用于根据所述有效载荷信号生成噪声信号;

电平检测器,用于检测所述原始音频信号的信号电平;

调制器,用于响应所述原始音频信号中所检测的信号电平的增加或降低而分别增加或降低所述噪声信号的电平,以便生成调制的噪声信号;

信号延迟设置,用于将延迟  $\tau$  施加到所述原始音频信号;以及

组合器,用于将延迟后的原始音频信号与所述调制的噪声信号组合;

其中所述调制器用于将小于所述延迟  $\tau$  的延迟施加于所述噪声信号的下降包络,使得所述噪声信号的电平中的降低相对于所述原始音频信号的信号电平中的相应降低是时间超前的。

2. 如权利要求 1 所述的设备,其特征在于,所述调制器用于将等于所述延迟  $\tau$  的延迟施加于所述噪声信号的上升包络,使得所述噪声信号的电平中的增加相对于所述原始音频信号的信号电平中的相应增加不是时间超前的。

3. 如权利要求 1 或 2 所述的设备,其特征在于,所述有效载荷信号是法务鉴定标记信号。

4. 如权利要求 1 或 2 所述的设备,其特征在于,所述噪声发生器是由一些或全部所述有效载荷信号播种的伪随机噪声发生器。

5. 如权利要求 4 所述的设备,其特征在于,所述噪声发生器可用于根据安全密钥对所述有效载荷数据的至少一些加密。

6. 如权利要求 1 或 2 所述的设备,其特征在于,所述组合器可用于将所述调制的噪声信号加到所述原始音频信号。

7. 如权利要求 1 或 2 所述的设备,包括两组或两组以上的电平检测器、调制器和组合器,设置为针对两个或两个以上的各自音频频带操作,所述设备包括用于将两个或两个以上的组合器的输出组合的部件。

8. 一种音频处理方法,其中将基于有效载荷信号的信号插入到原始音频信号中,所述方法包括如下步骤:

根据所述有效载荷信号生成噪声信号;

检测所述原始音频信号的信号电平;

响应所述原始音频信号中检测的信号电平的增加或降低而分别增加或降低所述噪声信号的电平,以便生成调制的噪声信号;

将延迟  $\tau$  施加到所述原始音频信号;以及

将延迟后的原始音频信号与所述调制的噪声信号组合;

其中增加或降低所述噪声信号的电平的步骤包括将小于所述延迟  $\tau$  的延迟施加于所述噪声信号的下降包络,以使得所述噪声信号的电平中的降低相对于所述原始音频信号的信号电平中的相应降低是时间超前的。

## 音频处理方法和装置

### 技术领域

[0001] 本发明涉及音频处理。

### 背景技术

[0002] 在诸如数字指纹处理或水印处理（可以通过术语法务鉴定标记（forensic marking）来通称）中，能以噪声模式的形式（如伪随机噪声信号）将有效载荷信号插入原始音频信号中。其目的一般在于噪声信号接近于不可感知，以及它在可以听到的情况下在主观上并不是干扰。这种类型的技术允许以无需改变原始音频信号的整体带宽、比特率和格式的方式添加各种类型的有效载荷。稍后可以通过相关技术恢复有效载荷数据，这样即使以多种方式在水印应用和水印恢复之间操纵或破坏了经水印处理的音频信号的情况下，有效载荷数据常常仍是有效的。

[0003] 可以添加的有效载荷数据的类型示例包括安全数据（例如用于识别盗版或非法拷贝的）、广播监视数据和描述原始音频信号表示的音频信号的元数据。

[0004] 在将噪声信号添加到原始音频信号之前可以对其调制。一般而言这意味着噪声信号的电平在原始音频信号的电平增加时增加，而在原始音频信号的电平降低时降低。由此，当噪声信号可被原始音频信号中的较响通道掩蔽时，可以包括更多有效载荷数据的噪声信号（提供有效载荷数据的潜在更佳恢复）。

[0005] 但是，如果噪声信号对原始音频信号跟随得太接近，则它可能变成可听的并潜在地在主观上构成干扰，尤其是有诸如鼓点等的声音时。

[0006] 在包络控制的音频处理系统中，可以对受控信号（在本示例中为噪声信号）的上升时间和下降时间应用一个时间常数。它们通称为增高和衰减（或释放）时间常数。如果将此测量值应用于本示例，则结果是原始音频信号电平的快速上升导致噪声信号中较慢的上升。这完全是可接受的 – 甚至在某些情况中是期望的。但是在原始音频信号电平中的突然降低会导致噪声信号电平中的较慢降低。在极端的情况中，这可能导致噪声信号瞬间地比原始音频信号大的非期望情况。

### 发明内容

[0007] 本发明提出音频处理设备，其中将有效载荷信号插入到原始音频信号中，该设备包括：

[0008] 噪声发生器，可用于根据有效载荷信号生成噪声信号；

[0009] 电平检测器，用于检测原始信号的信号电平；

[0010] 调制器，响应原始音频信号中检测到的信号电平的增加或降低来增加或降低噪声信号的电平，以便生成调制的噪声信号；

[0011] 用于将原始信号与调制的噪声信号组合的组合器；以及信号延迟装置；

[0012] 调制器相对于信号延迟装置来操作，使得噪声信号的电平中的降低相对于原始音频信号的信号电平中的相应降低是时间超前的。

[0013] 本发明通过如下方式来解决上文所描述的问题：提出一种时间超前释放功能，使得噪声信号的电平中的降低相对于原始音频信号的信号电平中的相应降低是时间超前的。换言之，相对于原始音频信号，噪声信号在原始音频信号开始下降之前就开始下降。可以关于系统中的任何释放时间常数和原始音频信号的音频带宽来设置此时间超前的量，以使噪声信号从不大于原始音频信号或使它们之间的任何差在被视为可接受的限度内。

[0014] 在所附权利要求中定义本发明的其他各个方面和特征。

#### 附图说明

[0015] 现在将参考附图仅以示例的方式描述本发明的实施例，其中：

[0016] 图 1 以示意方式图示包括指纹编码器的数字影院设置；

[0017] 图 2 以示意方式图示指纹检测器；

[0018] 图 3 是指纹编码器的操作的示意略图；

[0019] 图 4 以示意方式图示有效载荷发生器；

[0020] 图 5 以示意方式图示指纹流发生器；

[0021] 图 6 以示意方式图示频谱分析仪；

[0022] 图 7 以示意方式图示频谱跟随器；

[0023] 图 8 至图 11 以示意方式图示包络跟随器的操作；

[0024] 图 12 是指纹检测器的操作的示意略图；

[0025] 图 13 是示出时间对准单元的一部分操作的示意图流程图。

[0026] 图 14 以示意方式图示划分成块的可疑材料和代理材料；

[0027] 图 15 以示意方式图示低通滤波器设置；

[0028] 图 16 以示意方式图示阈值限定的信号；

[0029] 图 17 以示意方式图示相关操作；

[0030] 图 18 以示意方式图示功率曲线；

[0031] 图 19 以示意方式图示去卷积训练操作；

[0032] 图 20 以示意方式图示幅度曲线；

[0033] 图 21 以示意方式图示阈值和内插的幅度曲线；

[0034] 图 22 以示意方式图示图 19 中所示的处理的中间结果；

[0035] 图 23 以示意方式图示脉冲响应；

[0036] 图 24 以示意方式图示平滑曲线；

[0037] 图 25 以示意方式图示平滑的脉冲响应；以及

[0038] 图 26 以示意方式图示数据处理设备。

[0039] 具体实施方式

[0040] 已经提出指纹处理或水印处理技术 - 更通常称为法务鉴定标记技术，这些技术适用于视频信号。例如参阅 EP-A-1 324 262。虽然通用数学框架在原理上可应用于音频信号，但是存在多个显著的技术差异。在本文描述中，“指纹”和“水印”将用于指示材料的法务鉴定标记。

[0041] 要考虑的主要因素之一是应该如何将指纹数据编码到音频信号中。人耳在灵敏度和动态范围方面与人眼有非常大差异，这使得许多先前的商业指纹处理方案在主观收听

(“A/B”)测试中均失效。

[0042] 人耳能够听到 48kHz 采样率下少于一个样本的相位差,而且它在任一时间具有 9 个幅度等级的工作动态范围。据此原理,考虑一种适合的编码方法,将指纹数据编码为仅添加到媒体中的低电平噪声信号。

[0043] 噪声具有许多使之可有利于本任务的音质特性,尤其是当它处于低电平时人耳往往忽略它,而且它是一般为平静(模仿风、激流或海浪的自然声音)而非刺激的声音。噪声流的随机特征还意味着几乎没有可能以例如可能对视觉感知施加频闪(strobe)影响或下意识信息的恶意使用的方式干扰大脑功能。

[0044] 现在将描述这种技术的实施。

[0045] 数学基础

[0046] 考虑指纹有效载荷“矢量”(例如值流) $P = p[1] \dots p[n]$ 。

[0047] 对于嵌入过程,将该有效载荷添加到音频信号矢量(例如样本流) $V = v[1] \dots v[n]$ ,以得到经水印处理的有效载荷矢量  $W = V + P$ 。

[0048] 该有效载荷矢量  $P$  的元素是统计上独立的随机变量,平均值为 0,标准差为  $\alpha^2$ ,其中  $\alpha$  称为水印的强度,书写为  $N(0, \alpha^2)$ 。简单地说,该符号用于指示该有效载荷是高斯随机噪声流。缩放噪声流,使得标准差作为音频信号在范围  $\pm 1.0$  内。这种缩放是重要的,因为如果未正确执行此操作的话,下面计算的相似度指标(“SimVal”)将不正确。注意这里的约定是将  $\pm 1.0$  视为音频域中的“满标度”,这样在本例中高斯噪声流的许多样本实际将大于满标度。

[0049] 对于提取过程,从经水印处理的可疑矢量(例如考察的音频材料的盗版拷贝) $W_s$  减去初始代理矢量  $V$ ,以得到可疑有效载荷矢量  $P_s = W_s - V$ 。换言之, $P_s = \text{可疑音频流} - \text{代理音频流}$ 。

[0050] 为了测试内容是否是利用候选有效载荷矢量  $P$  进行水印处理的,在候选有效载荷矢量  $P$  与归一化可疑有效载荷矢量  $P_s$  之间执行内循环相关(书写为“ $\cdot$ ”),以得到相似度值,下文表示为 SimVal:

[0051]  $\text{SimVal} = (P_s / |P_s|) \cdot P$

[0052] 其中  $|P_s|$  是  $P_s$  的矢量幅度,表示  $|P_s| = \sqrt{P_s \cdot P_s}$ 。这里,  $\sqrt{\phantom{x}}$  表示平方根函数。注意归一化矢量意味着在矢量内缩放值,以使它们加起来刚好幅度为 1。

[0053] 此公式表示  $P_s$  与  $P$  之间的统计相关度,最大值接近于矢量长度的平方根。我们假定如果 SimVal 大于特定阈值  $T$ ,则  $P_s$  中存在有效载荷  $P$ ,以及如果  $\text{SimVal} \leq T$ ,则它不存在。

[0054] 为了赋予 SimVal 的值一些统计含义,通过如下公式将  $T$  的值与误报的概率相关联:

[0055]  $T = \sqrt{2 \ln(M^2 / p \sqrt{2\pi}))}$

[0056] 其中  $p$  是误报的概率,  $\ln$  是自然对数,以及  $M$  是填充大小(即就给定音频内容发行的惟一有效载荷矢量的数量)。例如,如果要求错误的概率好于一亿分之一,填充大小是 1000,则值 SimVal 需要大于 8。

[0057] 一般来说,使用本技术,SimVal 为 10 是盗版音频材料的法务鉴定分析中有意义的目标。对于特别大的填充  $M$ ,值为 12 可能更适合。在实验测试中,发现如果在可疑音频材料

数秒的的分析内达到 8 的值,则在另外数秒内一般将达到 12 的值。

[0058] 图 1 以示意方式图示数字影院设置,其中安全播放设备 10 接收加密的音频/视频材料连同解密密钥。解密器 20 将该音频和视频材料解密。将解密的视频材料提供到投影机 30,以投影到屏幕 40 上。将解密的音频材料提供到指纹编码器 50,指纹编码器 50 按上文所描述的应用指纹。

[0059] 一般来说,指纹对于该材料、影院和重播实例可以是惟一的,这将允许对盗版追溯到电影的特定显示。

[0060] 将经指纹处理的音频信号传送到放大器 60,放大器 60 驱动公知影院音响配置中的音箱 70 和低音音箱 80。

[0061] 还可以将指纹处理应用于视频信息。可以使用公知的视频指纹处理装置(未示出)。

[0062] 优选地,播放设备是安全的,因为它没有可用于获取未经指纹处理的音频(或实际视频)的外部连接的密封的单元。当然,放大器 60 和投影机 30 无需一定构成该安全系统的一部分。

[0063] 如果以来自影院演出的材料制作非法拷贝,例如使用影院内的摄像机制作非法拷贝,则与该电影相关联的音频内容将具有其中所含的指纹编码器 50 编码的指纹信息。为了研究或合法原因而建立它,可以将材料的可疑拷贝连同原始(或“代理”)材料以及用于生成原始指纹的密钥一起提供到图 2 的指纹检测器 80。在其最简单的方面,指纹检测器 80 生成特定指纹存在于可疑材料中的概率。下文将对检测过程作更详细的描述。

[0064] 嵌入过程

[0065] 在视频指纹处理中,这些技术一般是基于帧的(帧为视频域中正常处理块大小),指纹有效载荷矢量整体(以低电平)埋藏在每个帧中。在一些系统中,在帧的“较忙”图像区域中,以及在不严重更改视频内容的特征的情况下难以或不可能移除的较低空间频率处,将指纹的强度设为较大。该理念在于,在许多帧上,可以累计每个帧的相关,就像对一个矢量执行相关一样;如果可疑有效载荷  $P_s$  与候选有效载荷  $P$  之间存在真实的统计相关,则相关会一个帧接一个帧地继续升高。

[0066] 对于音频来说,一般没有此类正常处理块。

[0067] 在本实施例中,出于快速傅里叶变换(FFT)运算效率的原因,将音频版本的处理块大小设为 2 的乘幂个音频样本,例如 64k 个样本(65536 个样本)。还要注意矢量长度与处理块是相同大小的。

[0068] 能以与视频系统相同的方式累计这些音频帧的连续相关。

[0069] 每个内容样本有有效载荷矢量的一个样本。有效载荷还集中在“中频”,因为可以完全丢失高频内容(即  $> 5\text{KHz}$ )和低频内容(即  $< 150\text{Hz}$ ),而不会有音频质量不可容许地丢失。这些频率的丢失可能是劣质录制设备或技术在盗版部分上留下的人为痕迹,或它们可能被盗版故意移除以试图抑制指纹恢复过程。因此更适合将有效载荷集中到更具主观重要性的中频,即如果不严重降低质量便不能容易地移除的频率。

[0070] 一般地说:

[0071] 1. 有效载荷播种基于 AES Rijndael 的伪随机数流以生成噪声流。

[0072] 2. 根据音频流的感知分析对噪声流“整形”。

[0073] 3. 在低电平将整形的噪声流加到音频流。

[0074] 生成的噪声流在其中包含多个层,每个层从有效载荷数据的不同子集生成。将认识到可以在有效载荷内包含其他数据,如帧编号和 / 或日期 / 时间。

[0075] 通过对移动计数器重复应用 256 位 Rijndael 加密以生成随机数流。然后将这些数缩放到  $\pm 1.0$ ,以产生满标度的白噪声。通过向点对应用 Box-Muller 变换来将白噪声流转变成高斯噪声。

[0076] 在本实施例中,对于噪声流有 16 个层。由有效载荷的前 16 位播种伪随机噪声发生器的第一层,由有效载荷的前 32 位播种第二层,并依此类推,直到通过整个 256 位有效载荷播种第 16 层为止。

[0077] 感知分析包括简单的频谱分析,以便建立增益值来缩放音频流中的每个样本的指纹噪声流。该理念在于音频流中较响部分将隐藏较响强度的指纹噪声。

[0078] 进一步扩展此概念,将(要隐藏指纹的)音频流的中频内容拆分成多个频带(即 8 或 12 个),其中优选地将频带均匀分布在对数频率刻度上(当然还可以使用任何频带划分)。例如,这意味着大约将频谱分成八组。然后分开处理每个频带,以产生各自的增益包络以用于调制指纹噪声流中对应频带的振幅。当在所有频带中使用包络调制时,结果是噪声流听起来非常像原始音频信号的“鬼影”再现。更重要的是,这种鬼影再现由于与内容相似,所以当添加到原始材料时,对人耳来说变成不可听,尽管是在相对较高的信号电平下添加的。例如,即使调制的噪声在相对于音频 -30dB(分贝)一样高的电平添加调制的噪声,它在主观上仍然几乎不可听。

[0079] 本实施例使用 2049 个样本脉冲响应核来实施“砖墙”(陡边响应)卷积带通滤波器,以分离每个频带中的信息。为速度在 FFT 域中进行卷积。使用卷积滤波器作为带通滤波器而不是递归滤波器的一个重要原因是可以使卷积滤波器具有独立于频率的固定延迟。这一点是重要的原因在于,当添加噪声流时,进行任何给定频带的噪声流的调制必须与原始内容的实际包络对准。如果这些滤波器具有取决于频率的延迟,则所导致的未对准难以校正,这可能导致噪声的感知度增加,以及导致与频率的相关值的可能变化。

[0080] 图 3 是指纹编码器(如图 1 的编码器 50)的操作的示意略图。有效载荷发生器 100 产生要编码为指纹的有效载荷数据。如上文所描述的,这可能包括多种内容和其他标识符,它可以很好地对于内容重放实例是唯一的。下文参考图 4 进一步描述有效载荷发生器。

[0081] 将有效载荷提供到指纹流发生器 110。如上文所描述,这基本上是使用基于编码密钥的 AES-Rijndael 编码的随机数发生器,它产生取决于从有效载荷发生器 100 提供的有效载荷的输出序列。下文参考图 5 进一步描述指纹流发生器。

[0082] 将(要向其提供指纹的)源材料提供到频谱分析仪 120。它分析一个或多个频带中的源材料的振幅或包络。频谱分析仪将包络信息提供到频谱跟随器 130。频谱跟随器根据来自频谱分析仪 120 的包络信息调制指纹流发生器 110 输出的噪声信号。下文参考图 6 进一步描述频谱分析仪,以及参考图 7 进一步描述频谱跟随器。

[0083] 频谱跟随器 130 的输出是噪声信号,该噪声信号处于显著低于源材料电平的电平但一般跟随源材料的包络。噪声信号由加法器 140 添加到源材料。因此加法器 140 的输出是经指纹处理的音频信号。

[0084] 在源材料路径中示意示出延迟单元 150。这指示频谱分析和包络确定可以在比传

递到加法器 140 的源材料版本在时间上超前的源材料版本执行。下面将进一步描述该时间超前的特征。

[0085] 图 4 以示意方式图示有效载荷生成器。如上文所描述的,它会采用诸如序号、位置标识符和位置私钥之类的多种标识数据,并生成作为种子提供到指纹流发生器 110 的有效载荷数据 160。位置私钥可以由加密装置 170 用于对位置标识符加密。由逻辑 180 对有效载荷数据的多种分量进行位对准,以便作为种子输出。

[0086] 图 5 以示意方式图示指纹流生成器 110。它从有效载荷发生器 100 接收种子数据 160,并接收由扩充逻辑 200 扩充到十六个不同密钥  $K-1 \dots K-16$  的密钥数据 190。

[0087] 可以可选地由加法器 210 向种子数据 160 添加帧编号。

[0088] 流发生器具有十六个 AES-Rijndael 数发生器 220...236。它们中的每一个均从密钥扩充逻辑 200 接收各自的密钥。还由来自种子数据 160 的各自位集对每一个数发生器播种。由种子数据 160 的前 16 位对数发生器 220 播种。由种子数据 160 的前 32 位对数发生器 221 播种,并依此类推。这种设置允许建立有效载荷的层次结构,这使得更易于通过如下方式在解码阶段搜索特定指纹:首先搜索前 16 位的所有可能值,然后搜索第 17 至 32 位的可能值(已知前 16 位)并依此类推。

[0089] 每个数发生器 220...236 的输出提供到高斯映射装置 240...256。它们采用数发生器的输出(实际是白噪声),并应用已知的映射过程以产生具有高斯分布的噪声。

[0090] 由加法器 260 添加来自映射逻辑 240...256 的每个实例的高斯噪声信号,以生成作为输出的噪声信号 270。

[0091] 图 6 以示意方式图示频谱分析仪 120。它接收作为输入的(要执行指纹处理的)源材料,并生成作为输出的包络信息 280。

[0092] 频谱分析仪包括一组八个(在本示例中)带通滤波器 290...297,每个滤波器对来自源材料的各自频带的频率滤波。这些滤波器可能在频率上重叠或不重叠,八个滤波器覆盖整个可用频率范围的程度可以是百分之百或常常远少于此。与八个滤波器相关的各自频带可以是连续的(即彼此相邻)或不连续的。所使用的滤波器(频带)的数量可以少于或多于八个。因此,将认识到本文的描述仅仅是这些滤波器可操作的方式的一个示例。

[0093] 在本例中,滤波器处理中频范围,从大约 150Hz 至大约 5kHz。这被划分成八个对数上相等的频带,因此每个频带延伸覆盖大约八组的其中一组。用于带通滤波器 290...297 的滤波技术是根据上文描述的。

[0094] 在每个带通滤波器的输出处是包络检测器 300...307。这生成与各自带通滤波器输出端的滤波的源材料的包络相关的包络信号。

[0095] 图 7 以示意方式图示频谱跟随器。频谱跟随器从频谱分析仪 120 接收包络信息 280,并从指纹流发生器 110 接收高斯噪声信号 270。

[0096] 将高斯噪声信号 270 提供到一组带通滤波器 310...317。将它们设置为具有与频谱分析仪 120 的对应滤波器 290...297 相同(或尽可能相近)的响应。这在噪声频谱内生成八个频带。将每个滤波的噪声频带提供到各自包络跟随器 320...327。这采用源材料中与该频带的包络相关的包络信号,并在相同频带中调制滤波的噪声信号。由加法器 330 将所有包络跟随器 320...327 的输出求和,以生成整形的噪声信号 340。

[0097] 包络跟随器可以包括缩放装置,使得最终整形的噪声信号 340 相对于源材料处于

适合的电平,例如相对于源材料的  $-30\text{dB}$ 。

[0098] 如上文所描述的,由加法器 140 将整形的噪声信号 340 添加到源材料,以生成作为输出信号的经指纹处理的源材料。

[0099] 可以分开或同步地对不同的音频声道(如左声道和右声道)执行指纹处理过程。但是优选的是对每个声道使用不同的噪声信号,以避免盗版者通过比较多个声道试图推算(然后移除或毁坏)指纹。在两种情况中,包络信号 280 优选地与指纹编码的单独音频声道相关。

[0100] 现在将参考图 8 至 11 更详细地解释上文所描述的包络检测和包络跟随。注意,在上文所描述的频谱跟随器的情况中,包络跟随是针对每个声道或频带来进行的。而且,可以取决于可应用于频带的音频频率或频率范围、例如取决于该频带内信号的最快上升时间来设置下文描述的时间常数。这允许它们作为一组通过仅更改时间常数与最快上升时间之间的关系来调整。

[0101] 在图 8 至 11 中,水平轴表示任意标度上的时间,实曲线表示与源材料相关的包络信号的示例(采用示意形式),以及虚线表示(采用示意形式)包络跟随器 320...327 应用的调制。

[0102] 在图 8 中,响应源材料的包络的突然上升,包络跟随器应用时间常数来限制噪声信号的上升时间。这由虚线的左边部分表示,该虚线在时间上滞后于实线的更垂直上升。此类时间常数常常称为“增高”时间常数。但是,在图 8 至 11 的所有附图中要注意,虽然噪声信号的上升速率受到限制,但是噪声信号开始上升的时间与包络信号开始上升的时间相同(只存在检测延迟导致的微小时间差)。相对于包络信号使噪声信号上升的开始得以延迟(或甚至利用上文所描述的时间超前装置使之超前)是可能的,但是看上去这样做的好处不大。具体来说,使噪声信号的上升得以延迟能限制可隐藏在上升信号之后的有用有效载荷,并且使噪声信号开始时间超前可以提供与参考图 8 的包络的下降沿描述的那些类似的可听人为痕迹。

[0103] 相似地,在源材料包络的下降沿处,下降点划线所示的噪声包络的降低也受到“衰减”时间常数的限制。遗憾的是,这意味着随时间从  $t_1$  至  $t_2$ ,噪声信号大于源材料信号,这样噪声可能在主观上干扰收听者。

[0104] 图 9 图示包络跟随音频效应处理器中常见的情况,由此定义“持续”周期 350,该周期使包络跟随信号(在本例中为噪声信号)的衰减的开始得以延迟。这使得上文所描述的情况甚至更糟,因为噪声信号现在在时间  $t_1$  和  $t_3$  之间大于源材料信号。因此,在实施例中不使用持续周期。

[0105] 下文将参考图 10 和图 11 描述解决此问题的措施。。

[0106] 在图 10 中,噪声信号开始下降的时间相对于源材料的包络下降的时间超前了超前时间 360。在此示例中,这意味着噪声信号截至时间  $t_1$  已衰减到不重要的电平。

[0107] 在图 11 中,如果稍微缩减超前周期 360,则噪声信号在源材料的包络下降之前开始下降,但是截至时间  $t_1$  它尚未完成下降。这意味着在时间  $t_1$  和  $t_4$  之间仍有少量噪声存在,但是问题比图 8 所示的要小很多。

[0108] 因此,通过在比促使噪声下降的源材料包络的下降早的时间开始噪声信号的下降,可以降低或避免图 8 和 9 所示的主观上干扰的过度噪声。

[0109] 为了实现此目的,有必要在系统内某个位置包括延迟,以便可以在加法器 140 将源材料加到噪声的时间超前关系获取源材料的包络信息。图 3 所示的延迟是如何实现此目的的一个非常简明的示例。本领域技术人员将认识到许多其他可能性也是存在的。在上文示例中,将延迟加于从源材料到组合器 140 的路径中。在此设置中,频谱分析仪 130 可以按如下方式操作(如果得出超出一个,则针对每个包络信号):(a) 对于上升包络,(由延迟单元,未示出)对包络信号应用等于延迟组件应用的延迟  $\tau$  的延迟;以及(b) 对于下降包络,对该包络信号应用小于延迟  $\tau$  的延迟。

[0110] 提取过程

[0111] 指纹提取的主要阶段如下所示:

[0112] 1. 处理可疑材料,以尝试反转任何损坏或失真。

[0113] 2. 从可疑内容中减去所谓的代理内容(用于描述内容的未经水印处理的版本的术语)以留下可疑指纹。这依赖于能够在时间上将可疑材料与代理内容对准。在一些情况中,可以使用经水印处理的代理。当然代理中的水印可能通过相关而被检测到,但是它不妨碍检测其他水印,并且可以被忽略。以此方式,可以将安全的副本发送到操作提取过程的合约的第三方。

[0114] 3. 根据代理内容的频谱分析对可疑指纹“去整形(unshaped)”。

[0115] 4. 对于此内容的填充中的每个候选有效载荷,在内容的相对较短部分上将候选有效载荷与可疑有效载荷比较。如果值 SimVal 看上去是有希望的,则将该候选添加到将接受时间长很多的分析的候选短列表。

[0116] 图 12 是指纹检测器(如图 2 的检测器 80)的操作的示意略图。该检测器接收可疑材料、例如怀疑是盗版拷贝的一段内容和相同材料的明文(未经水印处理的)副本的所谓的代理材料。

[0117] 首先将可疑材料提供到时间对准单元 400。下文将参考图 13 至 18 描述此操作。但是简要地说,时间对准单元检测代理材料与可疑材料之间的任何时间偏移,由此使这两组材料在时间上能够对准。潜在地可以由时间对准 400 实现的对准要处于某个容差内,如  $\pm$  一个样本的容差。由去卷积器 410 执行进一步时间校正,以使两个信号之间能够完全对准,下文将对此予以描述。

[0118] 去卷积器对可疑材料应用脉冲响应,以尝试使它更像代理材料。这里的目的是要(至少部分)反转可疑材料中信号劣化的影响;下文列出了此类劣化的示例。

[0119] 为此,通过去卷积器训练单元 420 “训练”去卷积器 410。下文将参考图 19 至 25 描述去卷积器训练单元的操作,但是简要地说,去卷积器训练单元将时间对准的可疑材料与代理材料比较,以便推算出表示对代理材料进行了什么操作才将其转变成可疑材料的变换响应。由去卷积器 410 “逆向”应用这种变换响应。优选地,在可疑材料内的不同位置更新变换响应,以便表示该特定点上存在的劣化。在下文要描述的实施例中,去卷积器训练单元检测到的变换响应基于对可疑材料与代理材料的块的最近部分的预定成员检测到的响应的移动平均值。

[0120] 可以提供延迟 430,以补偿去卷积器和去卷积器训练操作。

[0121] 交叉归一化单元 440 然后执行操作以将去卷积的可疑材料和代理材料的幅度归一化。图 12 中所示,此操作对可疑材料执行,但是将认识到可以调整代理材料的幅度,或可

以同时调整二者的幅度。

[0122] 归一化之后,减法器 450 建立归一化且去卷积的可疑材料与代理材料之间的差值。将此差值信号传递到“去整形器”460,去整形器 460 设置为反转频谱跟随器 130 执行的噪声整形的结果。为此,对代理材料执行频谱分析阶段 470,此操作采用与图 3 的频谱分析仪 120 完全相同的方式执行。

[0123] 这样,可以认为频谱分析仪 470 和去整形器 460 采用与频谱分析仪 120 和频谱跟随器 130 相同的方式操作,所例外的是为了产生通常均匀的噪声包络作为去整形器 460 的输出而使用包络控制的增益值的倒数。将去整形器 460 生成的噪声信号  $P_s$  传递到比较器 480。按如下生成对比较器的其他输入  $P$ 。

[0124] 指纹发生器 490 采用与图 3 的有效载荷发生器 100 和指纹流发生器 110 相同的方式操作。因此,这里将不详述这些操作。指纹发生器 490 进而执行操作以产生可疑材料中可能存在的指纹的所有可能变体。进而测试每一个以推算各自的可能值  $SimVal$ 。

[0125] 当然,可能采用多个指纹发生器 490 和使用并行操作的多个比较器 480,以便将噪声流  $P_s$  一次与多于一个指纹比较。

[0126] 提供延迟 500、510 以补偿应用于可疑材料的处理延迟,以便使指纹发生器 490 生成的指纹适合地与可疑材料内可能包含的指纹时间对准。

[0127] 存储去整形器的输出是可能的,这样可以执行一个或多个与各自不同指纹(如模块 490、500、510 所处理的)的进一步比较,而无需重复产生去整形器 460 的输出的处理。

[0128] 时间对准

[0129] 处理可疑盗版信号的第一件事是发现与代理信号的真实同步。

[0130] 可以包括子样本延迟,以便在需要的情况下允许对再采样或 MP3 编码效应施加的任何子样本延迟/超前进行补偿。

[0131] 图 13 是示出时间对准单元 400 的一部分操作的示意图。该流程图的每个步骤由时间对准单元 400 的各自部件或功能实施。

[0132] 虽然在理论上来说通过(单个)直接相关过程对准可疑材料和代理材料是可能的,但是在例如电影音轨的基本材料的情况中,所需的相关处理是数量庞大的,因为这些处理操作通常随所涉及的音频样本数量的平方增加。因此,本过程旨在提供至少大致对准,而无需两个信号的完全相关。

[0133] 参考图 13,在步骤 600,将两个音频信号分成多个部分或块。对于这两个信号的每一个来说,这些块在大小上是相等的,但是无需是预定的大小。这样,一种选择是具有(比如说)64k 个样本的固定大小,而另一种选择是具有块的固定数量,使得将两段材料的较长一段(一般为代理材料)的总长度除以块的预定数量,以便达到时间对准处理的该特定实例的所要求的块大小。在任何情况中,块大小应该是至少两个样本。

[0134] 可以在图 13 的步骤 600 之前包括低通预滤波阶段(未示出)。就块大小而言,这可以减少两个信号之间任意未对准导致的任何人为痕迹。

[0135] 在步骤 605,为每个块建立每个信号的绝对值和检测的最大功率(参考该绝对值)。当然,也可以代之以建立不同功率特征,例如平均功率。其目的在于根据每一代理和可疑信号以功率特征信号结束,从而每个块具有小数量(例如 1 或 2 个)的值。本示例中每个块具有一个值。

[0136] 在步骤 610,对两个功率特征信号低通滤波或平滑处理。

[0137] 图 14 以示意方式图示将这两个信号划分成块,由此在本示例中代理材料表示电影的全部长度,可疑材料表示从电影截取的一段。

[0138] 图 15 以示意方式图示分开应用于两个功率特征信号的低通滤波器。在乘法器 611 将每个样本乘以一个系数,然后在加法器 612 加上加法器的输出和第二系数的积。这在乘法器 613 上进行。该过程产生每个信号的低通滤波的版本。

[0139] 在此阶段,这两个功率特征信号具有一般在 0 与 1 之间的幅度。滤波过程可能引入了一些微小的超过 1 的偏移,但是因为步骤 605 中绝对值检测的原因,没有零以下的偏移。

[0140] 在步骤 630,应用阈值。图 16 以示意方式图示了此情况。此类阈值的示例可以是 0.3,当然还可以使用多种其他值。

[0141] 按如下应用阈值。

[0142] 其目的在于将对应于该阈值的功率特征信号值映射到 1 的修正值。落在该阈值以下的任何信号值将被映射到 0 与 1 之间的信号值。落在该阈值之上的任何信号值将被映射到大于 1 的信号值。这样,实现此目的的一个简捷方式是将整个功率特征信号乘以  $1/\text{阈值的值}$ ,即在本例中是 3.33...

[0143] 之所以这是相关的,是因为下一步 640 将对这些信号应用幂率。这里的一个示例是,对每个信号求平方,就是说每个样本值自乘。但是,还可以使用大于 1、整数或非整数的其他乘幂。步骤 630 和 640 总的结果是,强化较高的信号值并弱化较低信号值的影响。此情况的产生是因为 0 与 1 之间的任何数自乘大于 1 次(例如求平方)变得更小,而大于 1 的任何信号值自乘大于 1 次变得更大。

[0144] 应用了幂率之后,在步骤 650 对结果信号执行可选的高通滤波过程。在步骤 660,减去每个信号的平均值,以便生成具有零平均值的信号。(此步骤对于更好地操作下面的相关步骤 670 有用)。

[0145] 最后,在步骤 670,对这些功率特征信号执行相关过程。图 17 中以示意方式图示此情况,其中对来自可疑材料的功率值填充 0,以便提供与代理材料相同长度的数据集。相关过程将(有希望)生成峰值相关,它与中心位置 702 的偏移 701 表示这两个文件之间的时间偏移。可以通过对代理或可疑信号应用相对延迟来校正此偏移。

[0146] 可以利用更小的块大小以及执行相关有关的限制范围来重复参考图 13 至 17 描述的该过程(取来自第一阶段的偏移 701 作为起始位置和近似答案)。实际上,该过程可以在适当降低块大小的情况下重复执行多于两次。要产生好处,块大小应该保持至少两个样本。

[0147] 图 18 以示意方式图示步骤 605 生成的功率特征信号和步骤 660 生成的滤波的功率特征信号。这里,阈值是 0.3,步骤 640 中的功率因数是 1.5 和应用了  $1/10$  的缩放。

[0148] 损坏反转

[0149] 损坏反转的目的是,以使盗版的内容变成尽可能接近于初始代理版本的方式变换盗版的内容。这样从盗版的版本减去代理所得到的可疑有效载荷  $P_s$  将尽可能小,这通常应该导致较大的 SimVal 值。

[0150] 对于音频来说,存在盗版者可能意外或故意施加的可能失真的长列表,每个失真潜在地导致 SimVal 值的降低:

- [0151] ■高、低、凹槽、频带或参数滤波
- [0152] ■压缩、扩展、限制、选通
- [0153] ■驱使过度、剪裁
- [0154] ■膨胀、闷音和其他声音增强效果
- [0155] ■重新采样、ADC 和 DAC 重新转换
- [0156] ■频率偏移、抖晃度、倒相、变速
- [0157] ■MP3- 系列有损编码 / 解码技术
- [0158] ■回音、回响、空间化

[0159] ■所谓的啞声消除 (de-essing)、杂音消除、爆音消除。

[0160] 为了应对尽可能多的这些损坏,指纹恢复设置包括通用去卷积器,这可以参考训练代理信号以显著地降低 / 移除卷积滤波器的操作所产生的任何影响。在电信(用于移除采用许多不同路径通过系统的信号所施加的非期望的回音)中和归档的材料恢复项目(用于移除老化损坏、或移除不完善的录制设备的人为痕迹)中可以发现其他先前使用的去卷积器。

[0161] 简言之,通过将可疑盗版的音频材料和代理版本变换到 FFT 域中来训练去卷积器。将期望的信号(代理)的实 / 虚值除以(使用复数除法)实际信号(盗版版本)的实 / 虚值,以获得将实际响应变换到期望的响应的脉冲响应核的 FFT。对结果 FFT 平滑处理,然后与先前的实例取均值,以推算出表示最近该音频信号的一般变换的 FFT。然后将该 FFT 转成时间域脉冲响应核,用于作为卷积滤波器来应用(一种过程,包括将时间域信号旋转,并对它应用窗口同步函数、如“加重平均”窗口以减少锯齿影响)。

[0162] 训练良好的去卷积器在原理上可以将应用于盗版版本的非线性增益效应的影响降少到原来的 1/10,例如通过麦克风压缩电路来实现。在实验测试中,发现去卷积器能够将每个块的 SimVal 值从 15 增加到 40。

[0163] 图 19 以示意方式图示去卷积器训练单元 420 应用的去卷积器训练操作。

[0164] 该过程开始于对可疑材料(700)和代理材料(710)逐个块进行快速傅里叶变换(FFT),其中块大小可以是例如 64k 连续样本。除法器 720 将 FFT 的其中一个除以另一个。在本例中,因为期望生成将应用于可疑材料的变换响应,所以除法器执行操作来将代理 FFT 除以可疑 FFT。

[0165] 平均器 730 对来自除法器 720 的当前除法结果与存储在缓冲器 740 中的 n 个最近除法结果取均值。当然,也将最近的结果添加到缓冲器,并将最久存储的结果废弃。n 的一个示例是 5。当然存储原始 FFT,形成两个平均值(一个用于代理,一个用于可疑材料)并除上这些平均值也是可能的,但是这会增加存储需求。

[0166] 转换器然后将取均值的除法结果(是复数结果)转换成幅度和相位表示。

[0167] 逻辑 750 移除任何小幅度值。这里,虽然删除了该幅度值,但是对应相位值保留不动。逻辑 750 仅对幅度值操作。通过线性插值法将删除的小幅度值替代为来自最接近的周围非删除的幅度值插入的值。

[0168] 图 20 和 21 中以示意方式图示了该过程,其中图 20 以示意方式图示作为一组幅度值(未示出相位值)幅度 / 相位转换器 740 的输出。删除小于阈值  $T_{mag}$  的任何幅度值,通过线性插值法在最接近的非删除值之间生成替代值 751、752、753。

[0169] 在转换器 770 将结果幅度值转换回复数表示之前,通过低通滤波器 760 对结果幅度值进行平滑处理。然后应用逆向 FFT 780。这生成更像图 22 中所示的脉冲响应。为了达到适于对可疑材料去卷积的形式,将脉冲响应旋转窗口大小的一半,以便将两个半波瓣连接成如图 23 中所示的中心峰值。这由逻辑 790 来执行。

[0170] 但是,图 23 中所示的逻辑 790 的输出仍不是完全适于去卷积。这是因为此响应的旁瓣 791 横跨整个窗口。如果在去卷积器 410 中使用此类响应,则这可能导致锯齿问题。因此,调制器 800 将图 23 的响应乘以如图 24 所示的同步窗口函数,以便产生如图 25 所示的必需的脉冲响应。这才是提供到去卷积器 410 的脉冲响应。

[0171] 电平匹配

[0172] 在去卷积操作之后,使盗版信号尽可能地与代理信号的电平匹配。在实际中,试验测试显示执行此操作的一种有用方式是将两个信号的平均幅度匹配而非将其峰值匹配。

[0173] 一旦实现这三个步骤(时间对准、去卷积和电平匹配),则从盗版材料减去代理信号以留下可疑有效载荷  $P_s$ 。

[0174] 可疑有效载荷的提取

[0175] 注意,嵌入过程中由噪声整形器产生的有效载荷信号与进入其中的高斯噪声流非常不同。为了找出值 SimVal 而恢复更接近匹配候选有效载荷高斯噪声流(从统计意义上而言)的可疑有效载荷信号,恰当的是将噪声整形的结果反转,即对该有效载荷信号“去整形”。

[0176] “去整形”使用相同的噪声整形组件来实现,所例外的是并不将噪声流乘以增益值,而是应用除法。

[0177] 从技术角度来看,另一种可能的方法、即在比较之前对候选有效载荷流执行噪声整形是可能的,但是由于法律原因,它得不到支持。这是因为由统计上独立的样本组成候选流违背数字版权管理系统中采用的数学原理。对噪声流应用滤波器自动与样本相关。

[0178] 另一个原因是如果搜索的信号淹没在噪声里,则卷积技术往往操作更成功。在噪声中查找噪声流一般比在相似整形的余留音频信号中查找整形的信号更有效且更可靠(因为它得到更稳定的交叉相关)。

[0179] 最后,图 26 图示数据处理设备。这仅仅是作为可以如何实施图 1 的编码器 50 和图 2 的检测器 80 的一个实例来提出的。但是,应该注意至少在图 1 中,整个数字影院设置 10 优选地是一个没有外部连接的安全单元,所以至少指纹编码器可以更好地作为硬连线设备、例如作为一个或多个现场可编程门阵列(FPGA)或专用集成电路(ASIC)来实施,。

[0180] 参考图 26,数据处理设备包括中央处理单元 900、存储器 910(例如随机存取存储器、只读存储器、非易失性存储器等)、提供与例如显示器 930 和诸如键盘、鼠标或二者兼有的用户输入设备 945 的接口的用户接口控制器 920、诸如硬盘存储装置、光盘存储装置或二者兼有的存储装置 930、用于连接到局域网或因特网 950 的网络接口 940 以及信号接口 960。在图 26 中,以适用于指纹编码器 50 的方式示出信号接口,因为它接收未经指纹处理的材料并输出经指纹处理的材料。但是,当然可以使用该设备来实施指纹检测器。

[0181] 单元 900、910、940、920、930、960 是通过总线 970 来互连的。在操作中,由存储介质(例如光盘)或通过网络或因特网连接 950 来提供计算机程序,并将其存储在存储器 910 中。由 CPU 900 来执行连续指令,以执行与如上文所描述的指纹编码或检测相关的功能。

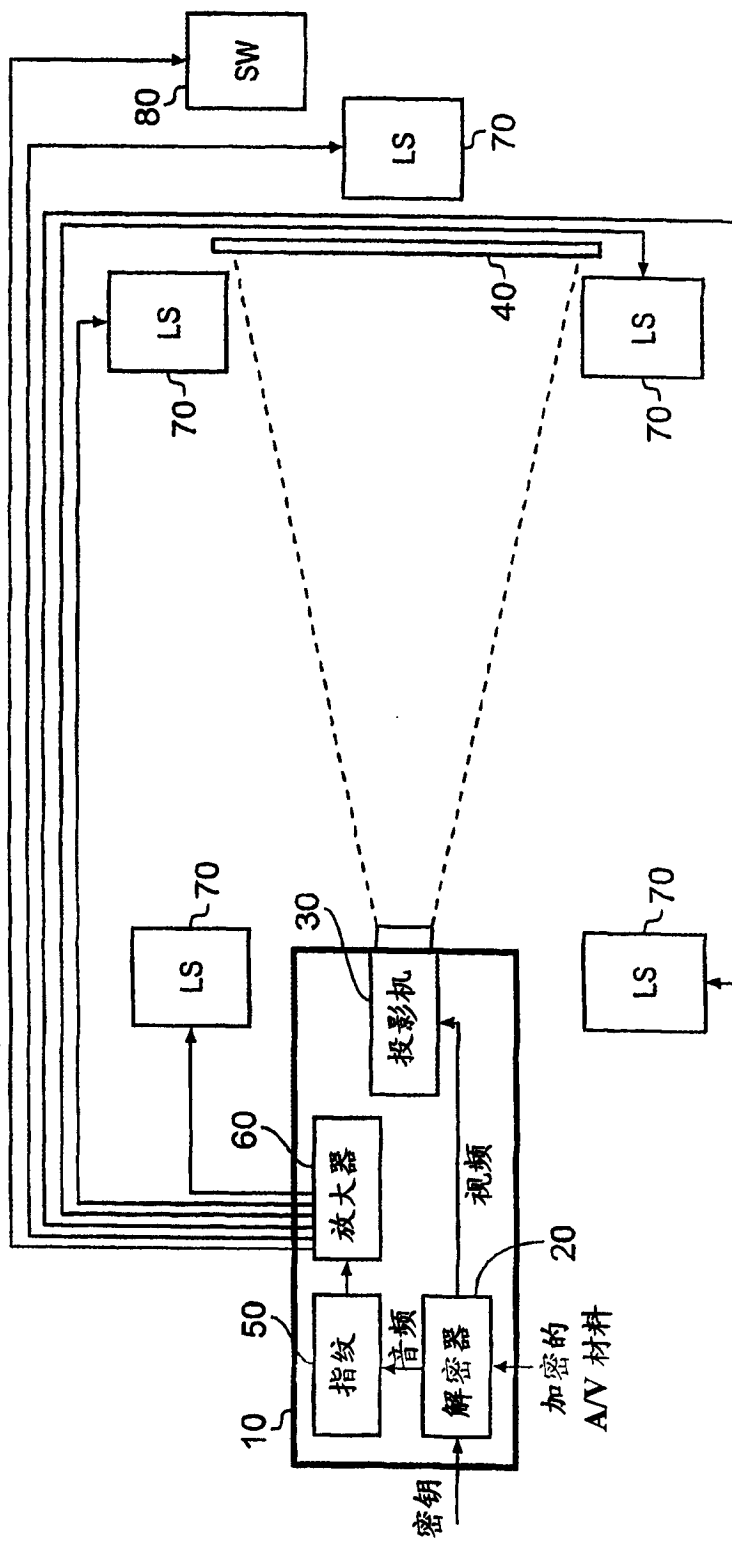
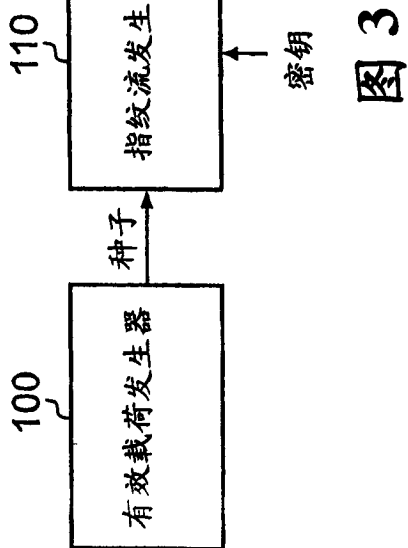
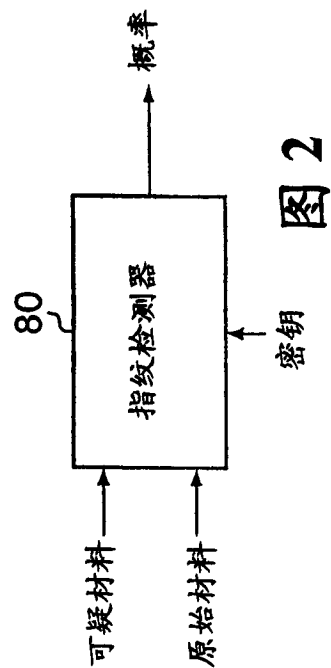
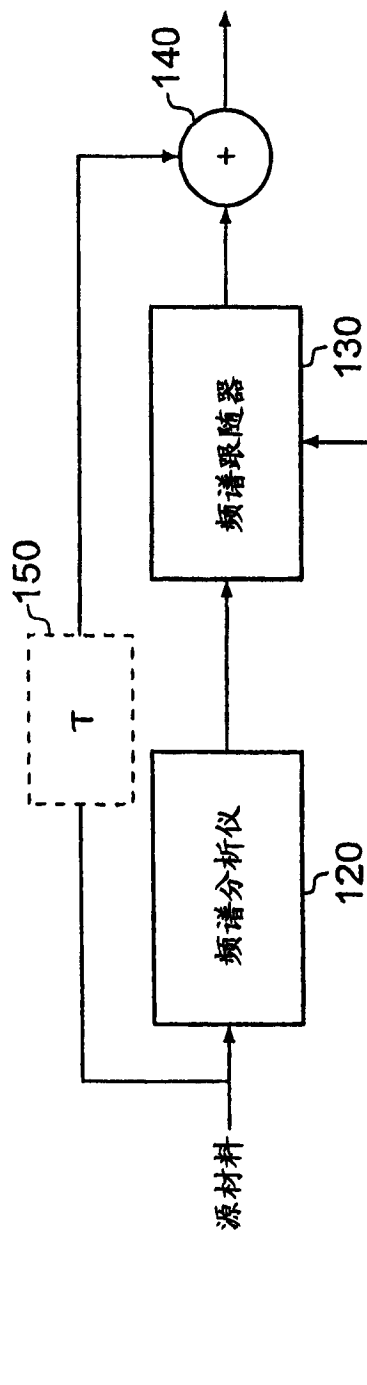


图 1



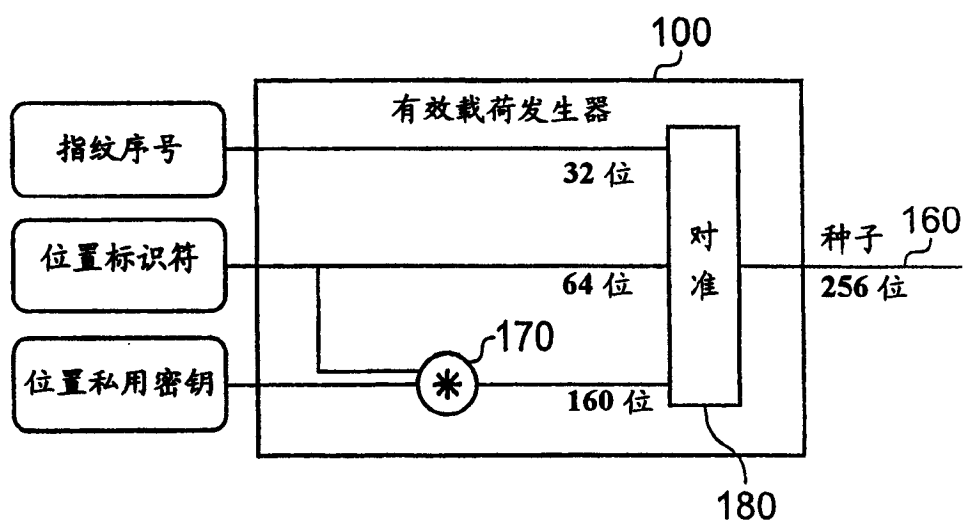


图 4

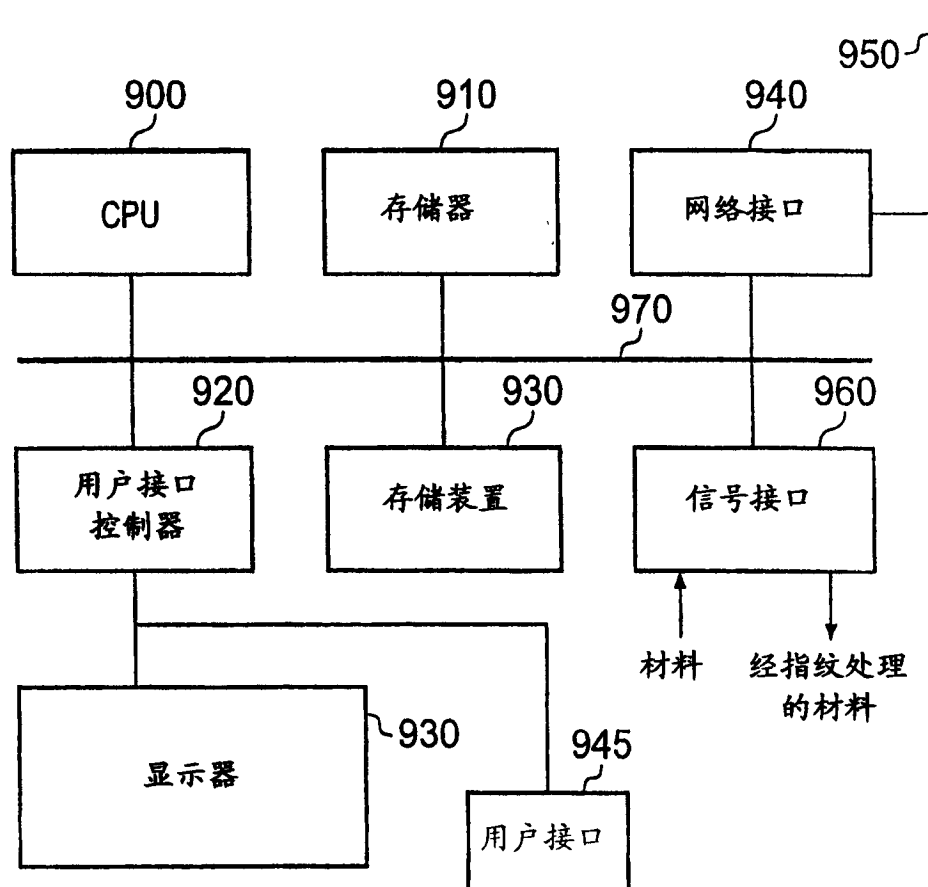
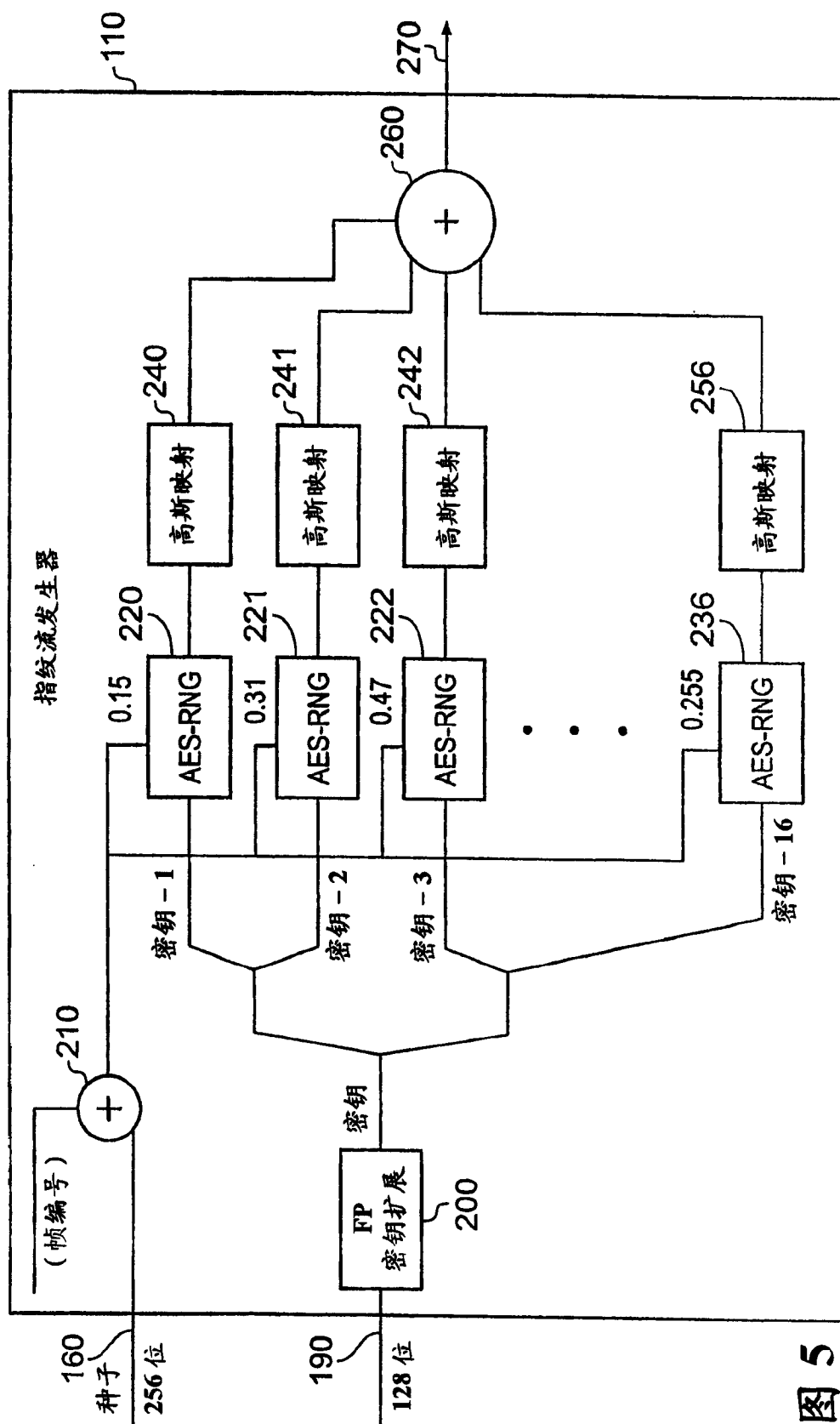


图 26



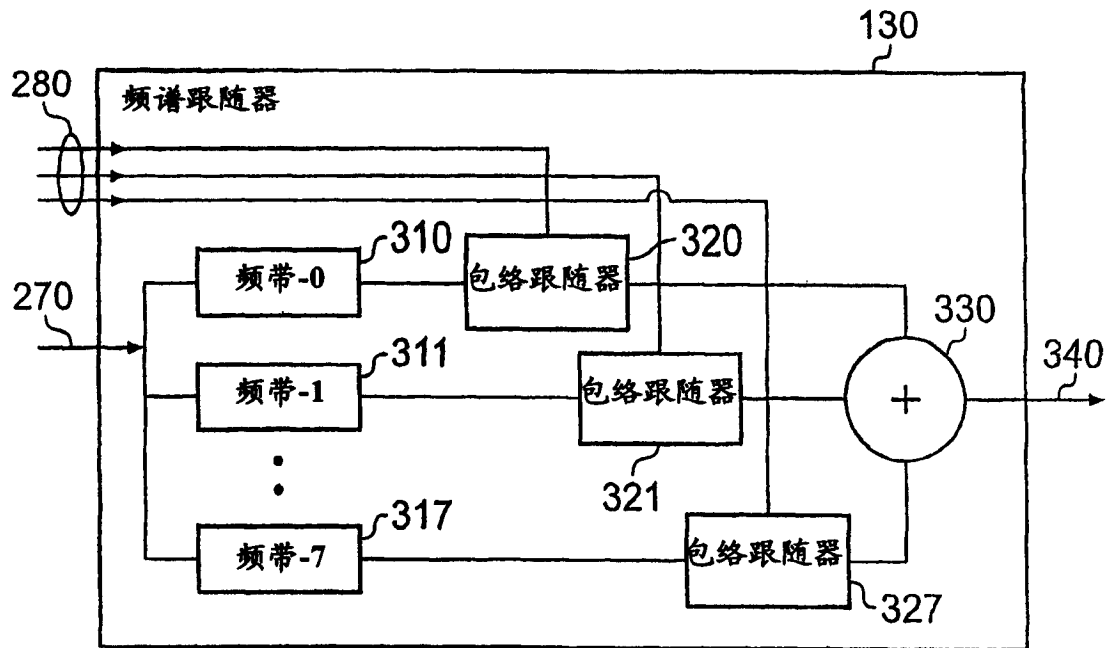


图 7

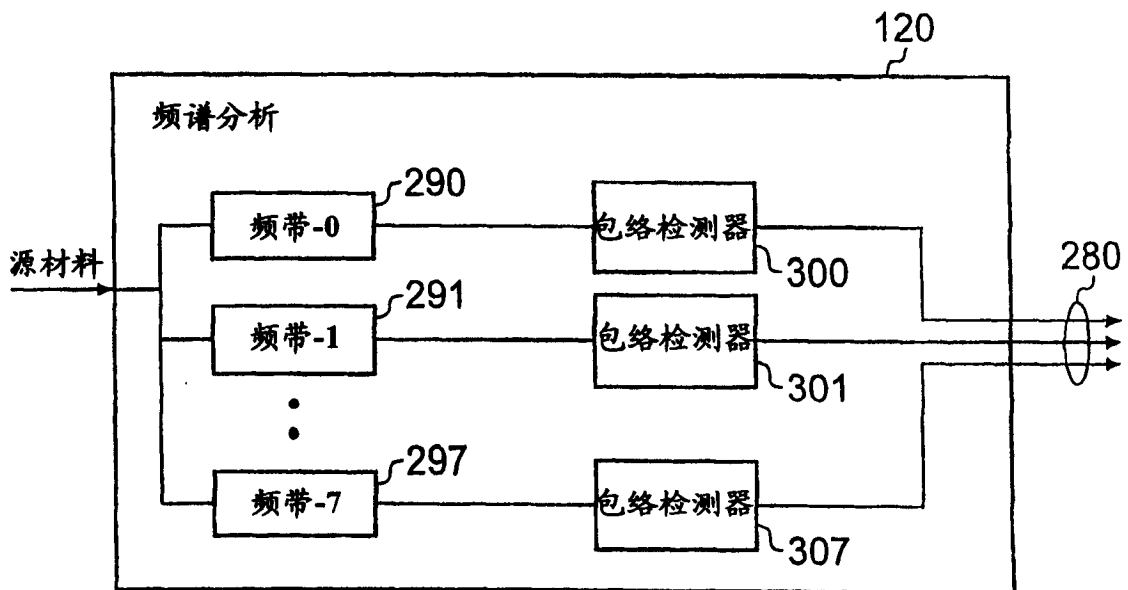


图 6

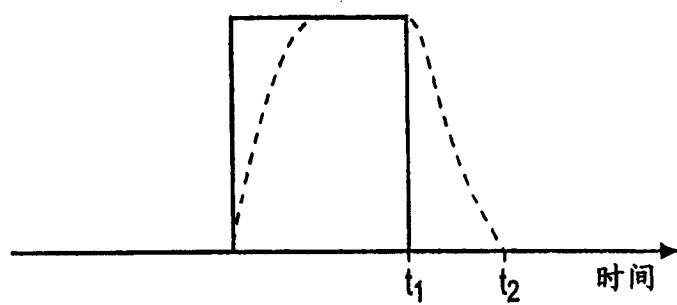


图 8

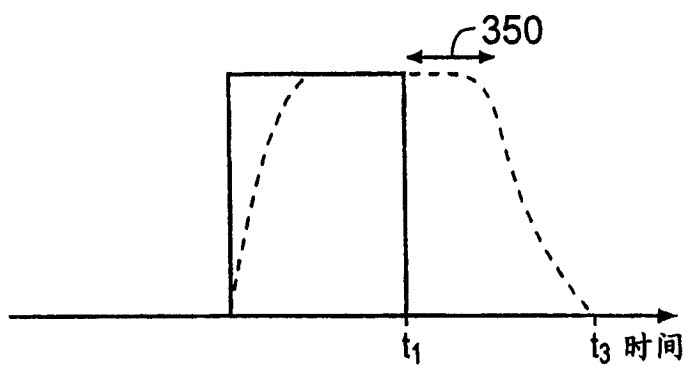


图 9

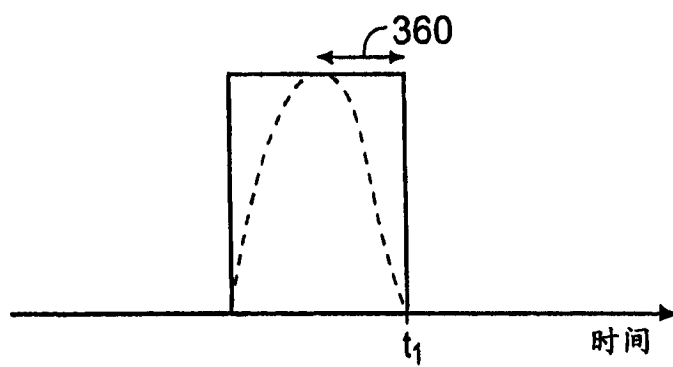


图 10

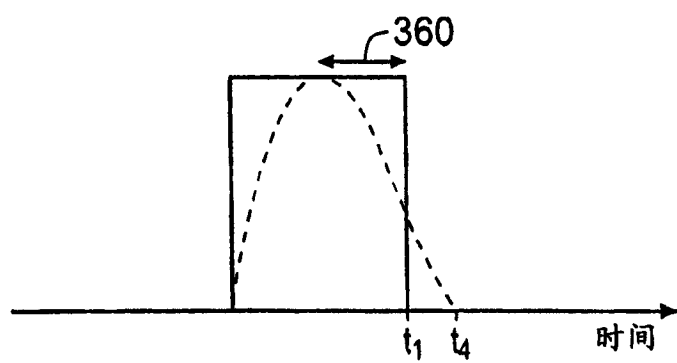


图 11

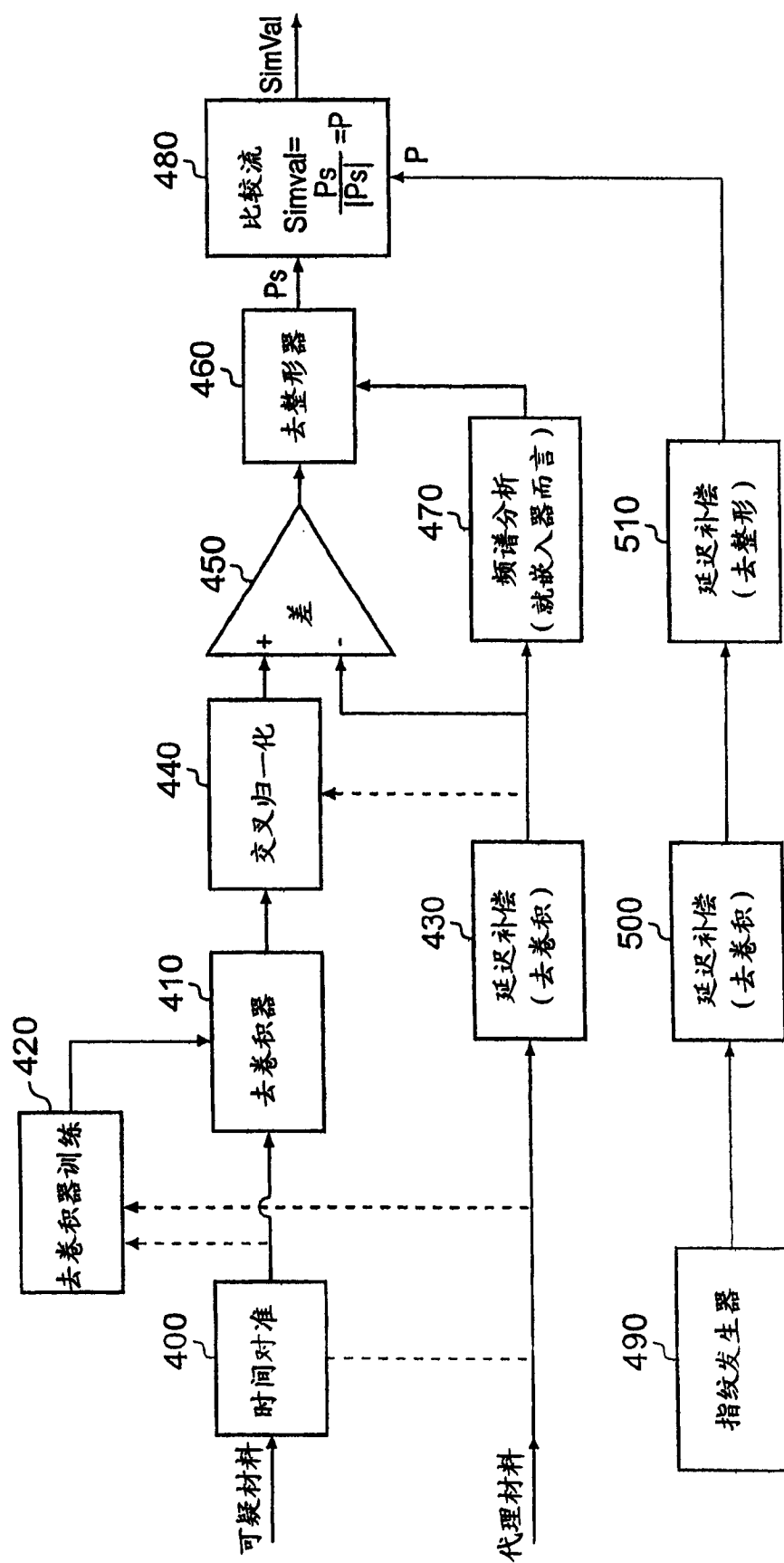


图 12

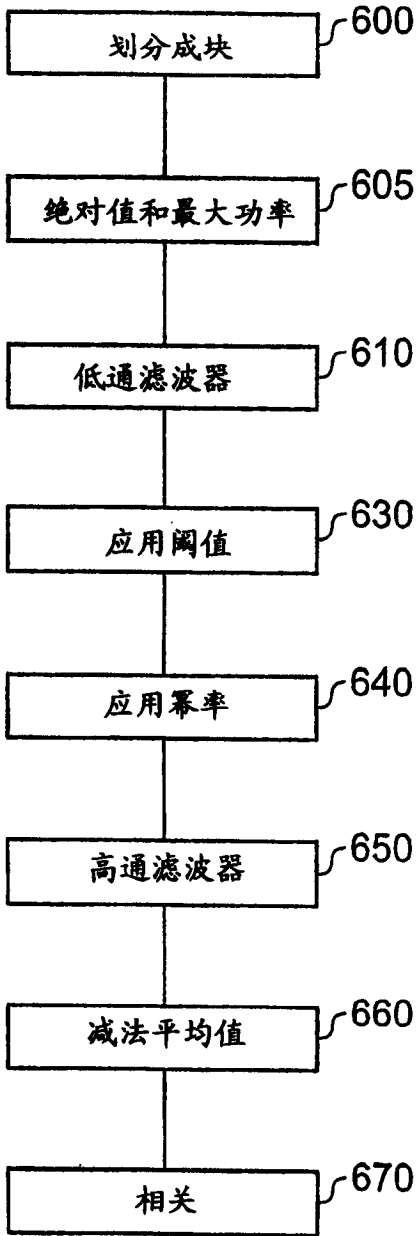


图 13

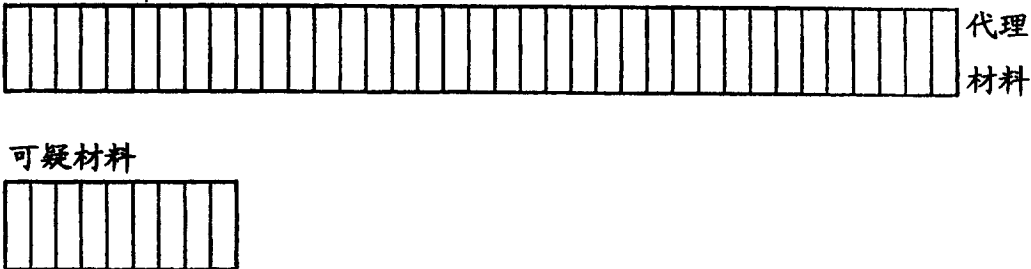


图 14

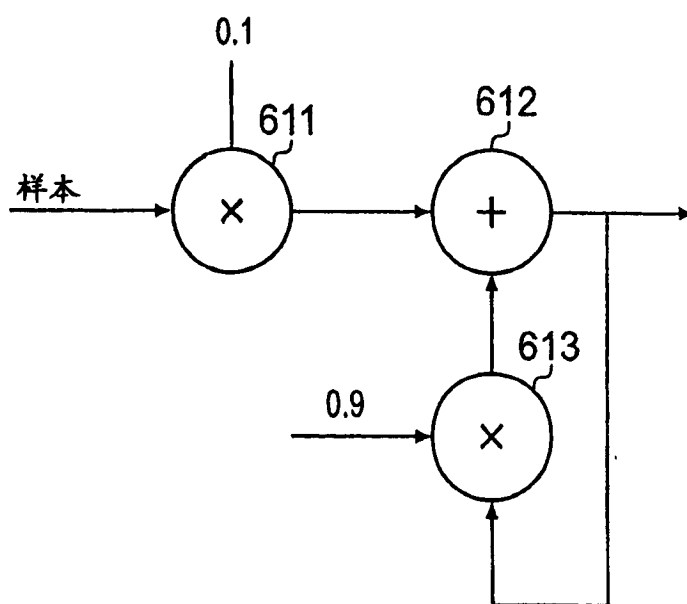


图 15

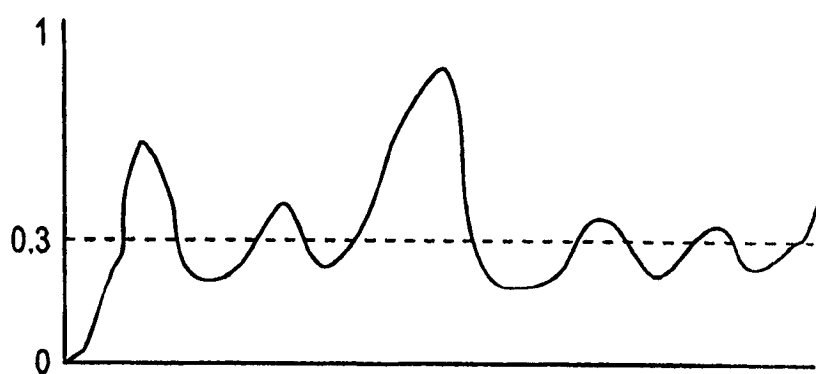


图 16

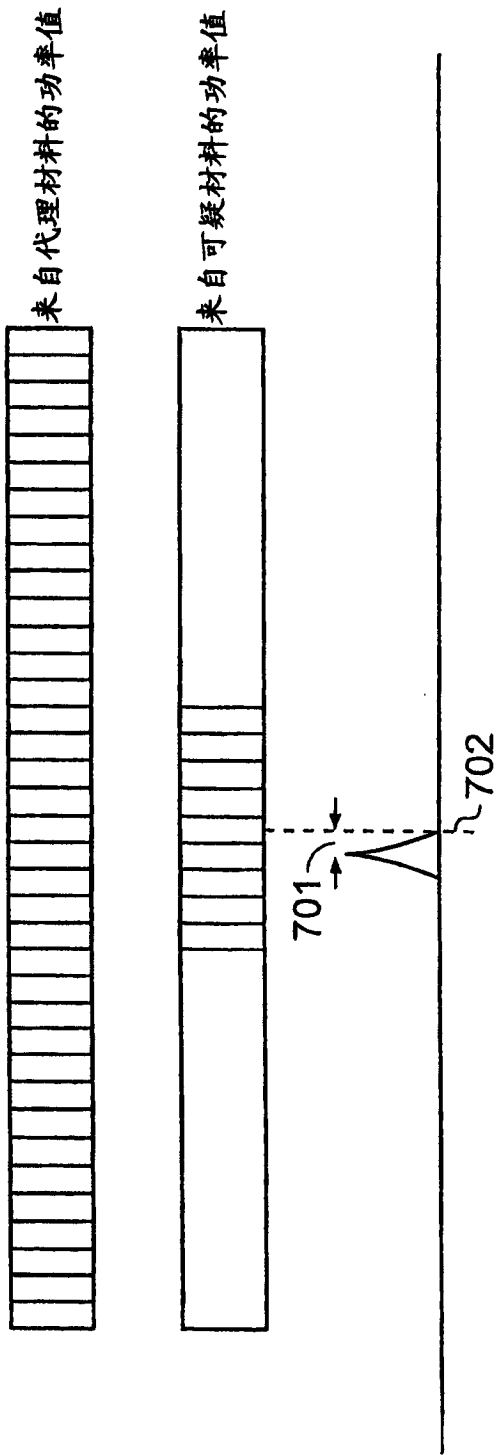


图 17

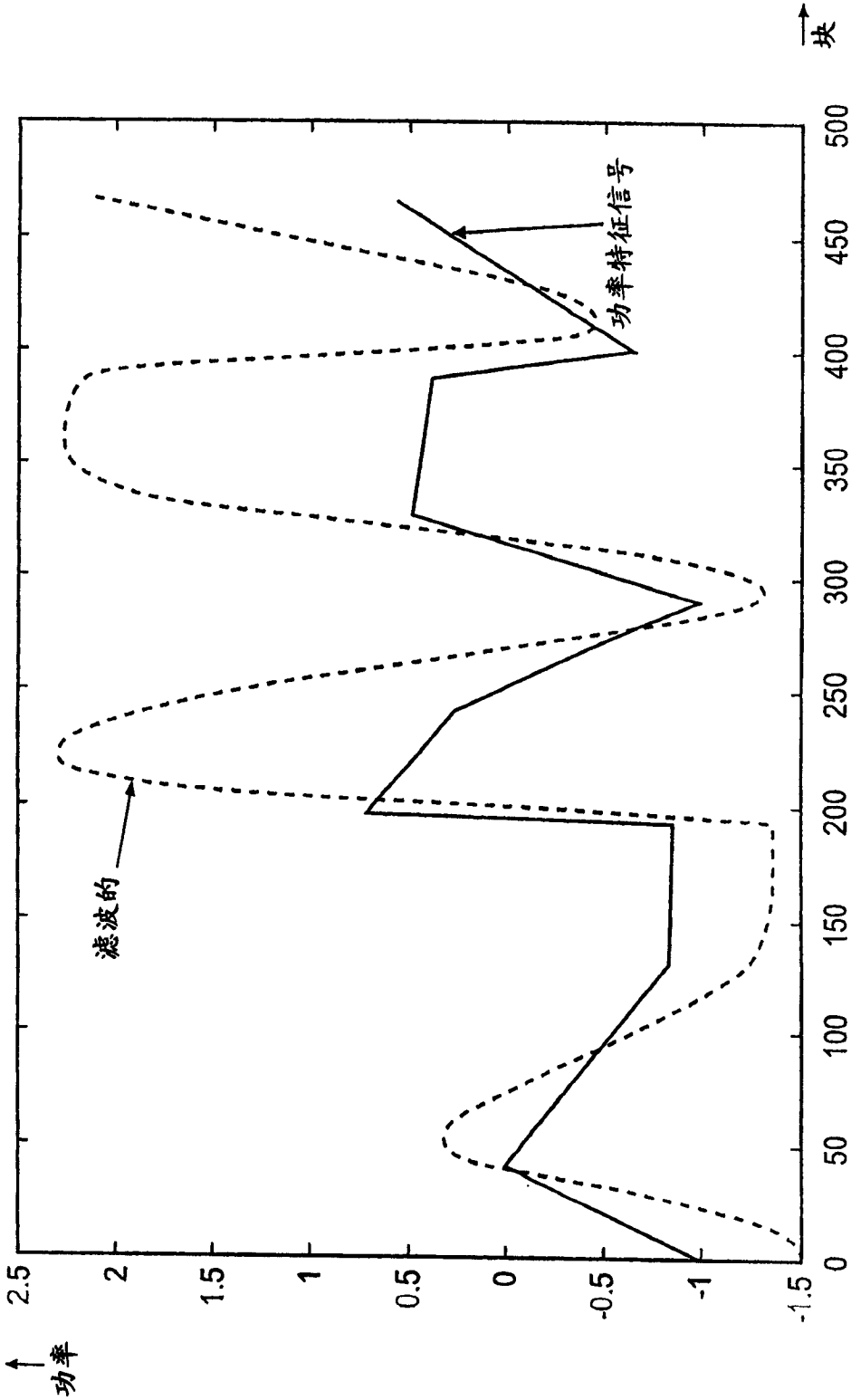


图 18

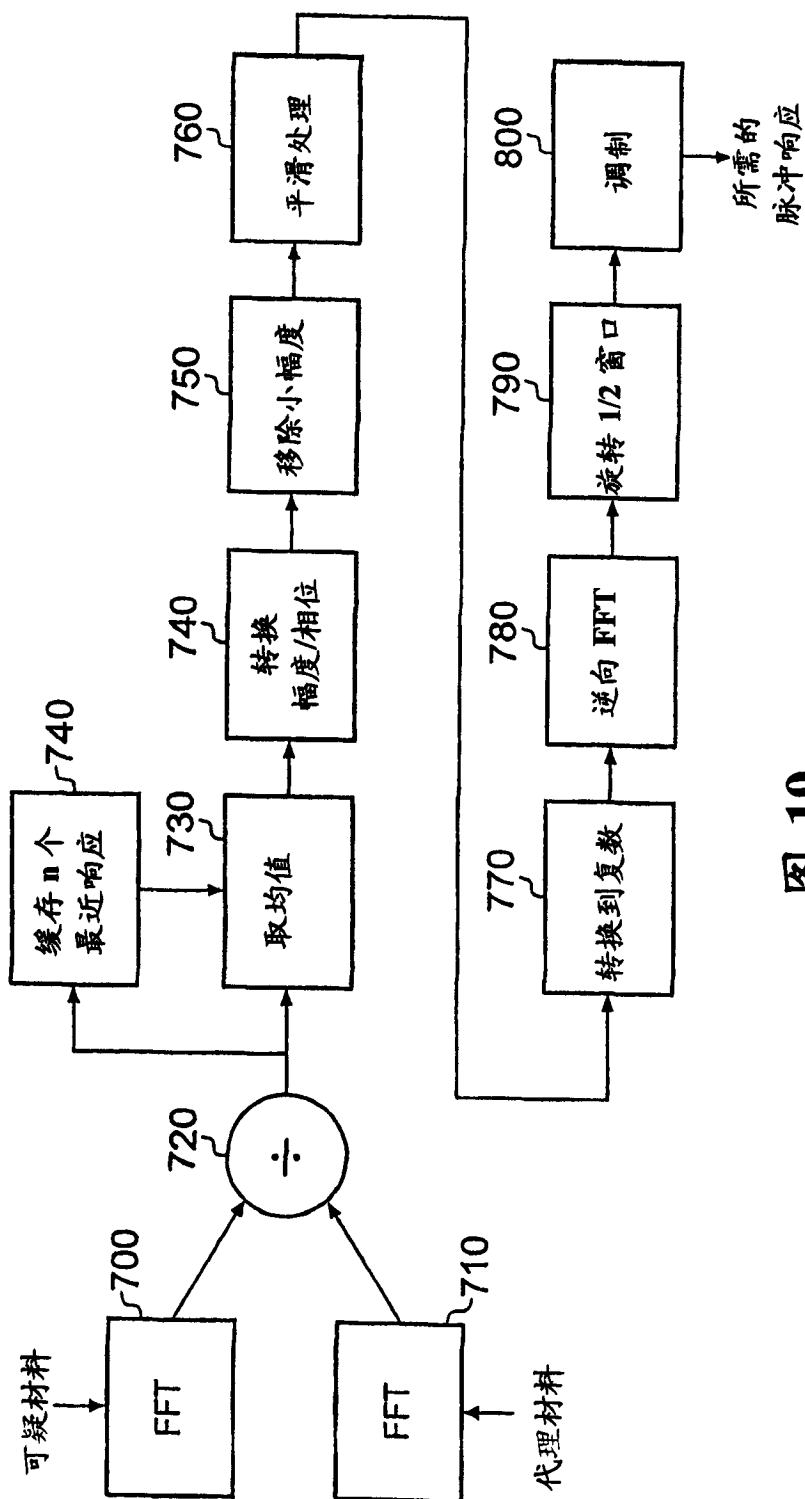


图 19



图 20

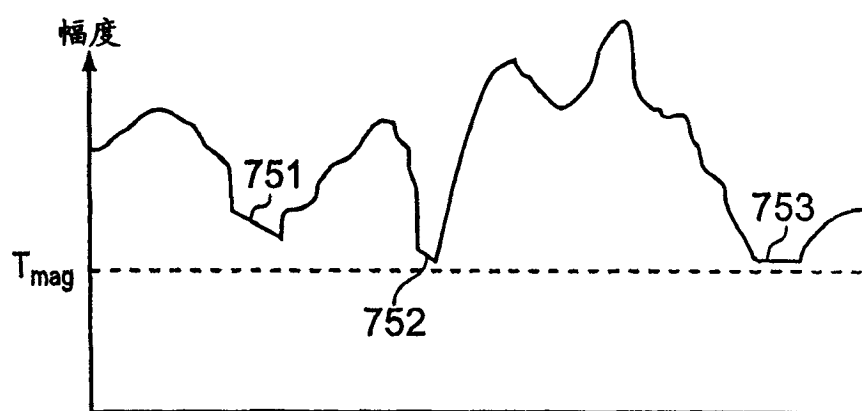


图 21



图 22

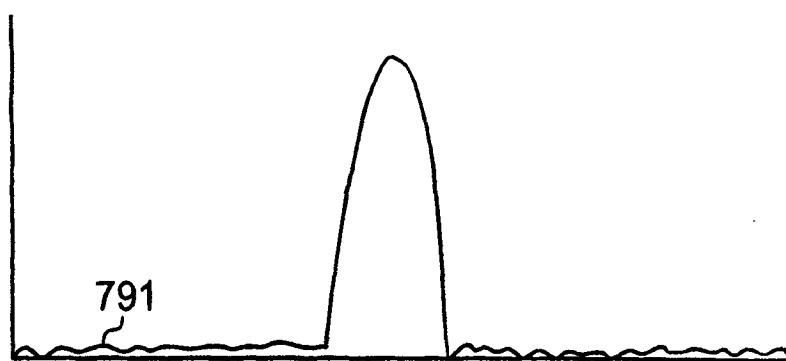


图 23

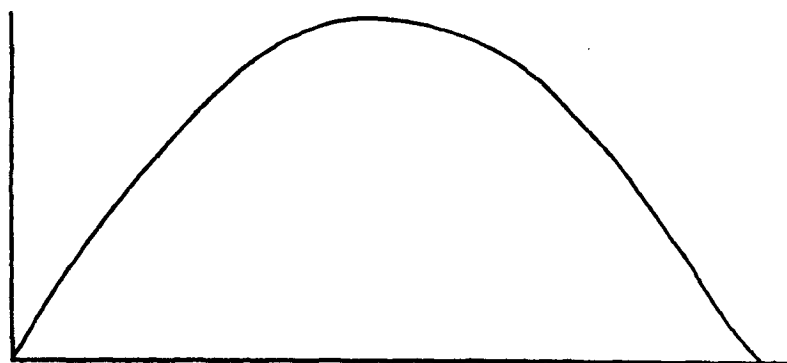


图 24

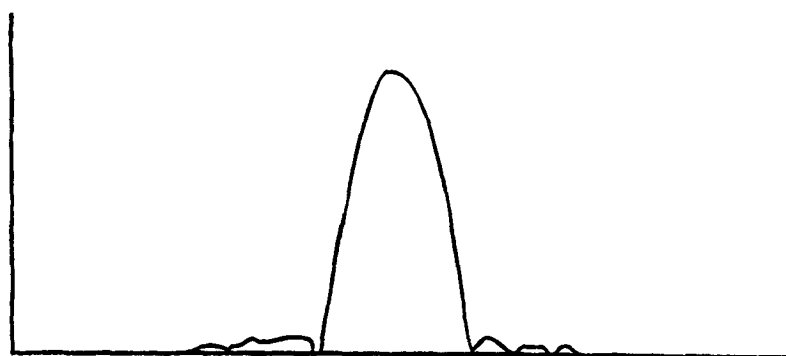


图 25