

(12) 发明专利

(10) 授权公告号 CN 101667420 B

(45) 授权公告日 2012. 08. 29

(21) 申请号 200910172047. 5

审查员 张鑫

(22) 申请日 2009. 09. 03

(30) 优先权数据

2008-226344 2008. 09. 03 JP

(73) 专利权人 索尼株式会社

地址 日本东京都

(72) 发明人 宫岛靖

(74) 专利代理机构 北京集佳知识产权代理有限公司

11227

代理人 陈炜 许伟群

(51) Int. Cl.

G10H 1/00 (2006. 01)

(56) 对比文件

US 20050273328 A1, 2005. 12. 08, 说明书第 25-26、33-45 段, 附图 3.

CN 101093661 A, 2007. 12. 26, 全文.

JP 2008164932 A, 2008. 07. 17, 全文.

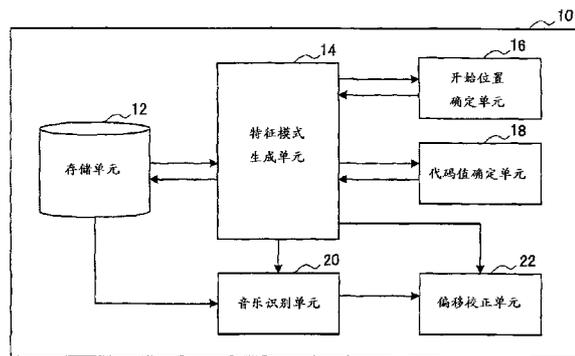
权利要求书 2 页 说明书 12 页 附图 21 页

(54) 发明名称

音乐处理方法及音乐处理装置

(57) 摘要

本发明涉及一种音乐处理方法、音乐处理装置和程序。本发明提供了一种包括以下步骤的音乐处理方法:根据音乐数据中在时间轴上从编码开始位置开始的每一定时段中的音量转变来确定代码值;以及使用在多个时段内确定的一系列代码值来生成表明所述音乐数据的音量转变特征的特征模式。代码值可以被配置成例如表明一定时段中的平均音量是从先前时段的平均音量增加还是减小。



1. 一种音乐处理方法,包括以下步骤:

根据音乐数据中在时间轴上从编码开始位置开始的每一定时段中的音量转变来确定代码值;

使用在多个时段内确定的一系列代码值来生成表明所述音乐数据的音量转变特征的特征模式;

通过比较针对所述音乐数据生成的特征模式与任意音乐数据的特征模式来计算所述音乐数据的特征模式的相似度;以及

基于与相似度高的两段音乐数据的特征模式对应的编码开始位置来校正音乐数据的偏移或者与所述音乐数据有关的元数据的偏移。

2. 根据权利要求 1 所述的音乐处理方法,其中:

所述代码值表明一定时段中的平均音量是从先前时段的平均音量增加还是减小。

3. 根据权利要求 1 所述的音乐处理方法,还包括以下步骤:

通过获取音乐数据中在时间轴上的窗口宽度中的平均音量第一次超过全部或者部分音乐数据的平均音量时的窗口位置,来确定编码开始位置。

4. 根据权利要求 1 所述的音乐处理方法,其中:

将所述相似度计算为两个特征模式之间的匹配代码值的位数与总位数之比。

5. 根据权利要求 1 所述的音乐处理方法,其中:

基于表明在两个特征模式之间代码值匹配的位和表明代码值不匹配的位在二项式分布中的出现概率来计算所述相似度。

6. 根据权利要求 1 所述的音乐处理方法,其中:

基于两个特征模式的代码值的连续匹配部分的最大长度来计算相似度。

7. 根据权利要求 3 所述的音乐处理方法,其中:

在确定编码开始位置的步骤中使用不同窗口宽度来确定多个编码开始位置;以及在生成特征模式的步骤中基于多个编码开始位置来生成针对一段音乐数据的多个特征模式。

8. 根据权利要求 7 所述的音乐处理方法,包括以下步骤:

通过比较针对所述音乐数据生成的各个特征模式与任意音乐数据的特征模式来计算所述各个特征模式的相似度;以及

基于与多个特征模式之中所计算的相似度最高的特征模式对应的编码开始位置来校正所述音乐数据的偏移或者与所述音乐数据有关的元数据的偏移。

9. 一种音乐处理方法,包括以下步骤:

获取音乐数据中在时间轴上的窗口宽度中的平均音量第一次超过全部或者部分音乐数据的平均音量时的窗口位置;以及

基于音乐数据的窗口位置来校正所述音乐数据的偏移或者与所述音乐数据有关的元数据的偏移。

10. 一种音乐处理装置,包括:

代码值确定单元,所述代码值确定单元根据音乐数据中在时间轴上从编码开始位置开始的每一定时段中的音量转变来确定代码值;

特征模式生成单元,所述特征模式生成单元使用由所述代码值确定单元在多个时段内

确定的一系列代码值来生成表明在音乐数据中的音量转变特征的特征模式；

音乐识别单元,所述音乐识别单元通过比较针对所述音乐数据生成的特征模式与任意音乐数据的特征模式来计算所述音乐数据的特征模式的相似度;以及

偏移校正单元,所述偏移校正单元基于与相似度高的两段音乐数据的特征模式对应的编码开始位置来校正音乐数据的偏移或者与所述音乐数据有关的元数据的偏移。

音乐处理方法及音乐处理装置

技术领域

[0001] 本发明涉及一种音乐处理方法、音乐处理装置和程序。

背景技术

[0002] 近年来由于信息处理装置的性能改进而已经以新的方式使用数字化音乐数据。新的使用方式之一是再混合 (remixing)。一般而言,再混合是指一种用于组合多段音乐的全部或者部分以组成新的整段音乐的方法。虽然再混合本身在以往已经进行过,但是对于没有专用设备的用户而言变得更容易使用 PC(个人计算机)来再混合音乐数据,从而使更多用户再混合数据并且更多再混合数据在用户之间交换。另外,其中与显示的画面或者幻灯片的场景关联地播放音乐的幻灯片放映是音乐数据的新用途之一。这样的幻灯片放映已经广泛地使用在活动和产品推广时的演示中。

[0003] 当在再混合或者幻灯片放映中使用音乐数据时,使用附加于音乐数据的元数据是有效的。元数据不仅包括比如标题和艺人姓名这样的静态数据,而且包括通过对音乐数据的时序分析而获得的数据,比如节奏、节拍位置、小节长度或者位置、和弦进展和旋律类型。当使用这样的时序元数据时,变得更易于从音乐数据提取特定小节或者调整节奏和节拍位置。

[0004] 另外,只是享受播放音乐的其他用户可以利用元数据。例如使用由元数据表明的音乐特征在海量存储器介质如硬盘或者闪存中存储的段数众多的音乐数据之中快速地搜索所需音乐数据是有效的。

[0005] 有鉴于此,已经开发用于生成与音乐数据有关的元数据或者支持使用元数据的技术。例如,日本专利申请特许公开 2007-248895 公开了一种元数据创建装置,其能够自动提取音乐数据中的节拍位置或者小节头部位置并且向用户提供该数据使得用户可以容易地校正该数据。日本专利申请特许公开 2008-164932 公开了一种音乐编辑装置,其能够通过使用应用于音乐的元数据如节拍位置来适当地调整音乐数据播放的定时。

发明内容

[0006] 然而,在不同环境中记录于介质中的音乐数据可能由于数据读取装置的差别或者编码参数的差别而即使在原始音乐相同时仍然造成从数据开始点到演奏开始点的偏移差异。例如,当与音乐数据分开地提供时序元数据时,偏移差异可能造成元数据没有与音乐数据匹配。当元数据没有与音乐数据匹配时,使用元数据的再混合、幻灯片放映或者音乐数据搜索可能并不能提供用户所期待的结果。

[0007] 已经鉴于上述问题而做出了本发明,并且希望提供一种无论偏移差异如何仍然使用音乐数据或者音乐数据的元数据的新型或者改进的音乐处理方法、音乐处理装置和程序。

[0008] 根据本发明的一个实施例,提供了一种包括以下步骤的音乐处理方法:根据音乐数据中在时间轴上从编码开始位置开始的每一定时段中的音量转变来确定代码值;并且使

用在多个时段上确定的系列代码值来生成表明音乐数据的音量转变特征的特征模式。

[0009] 代码值可以表明一定时段中的平均音量是从先前时段的平均音量增加还是减小。

[0010] 音乐处理方法还可以包括以下步骤：通过获取音乐数据中在时间轴上的窗口宽度中的平均音量第一次超过全部或者部分音乐数据的平均音量时的窗口位置来确定编码开始位置。

[0011] 音乐处理方法还可以包括以下步骤：通过比较针对音乐数据生成的特征模式与任意音乐数据的特征模式来计算音乐数据的特征模式的相似度。

[0012] 相似度可以计算为两个特征模式之间的匹配代码值的位数与总位数之比。

[0013] 可以基于表明在两个特征模式之间代码值匹配的位和表明代码值不匹配的位在二项式分布中的出现概率来计算相似度。

[0014] 可以基于两个特征模式的代码值的连续匹配部分的最大长度来计算相似度。

[0015] 音乐处理方法还可以包括以下步骤：基于与相似度高的一段音乐数据的特征模式对应的编码开始位置来校正音乐数据的偏移或者与音乐数据有关的元数据的偏移。

[0016] 在确定编码开始位置的步骤中，可以使用不同窗口宽度来确定多个编码开始位置，并且在生成特征模式的步骤中可以基于多个编码开始位置来生成一段音乐数据的多个特征模式。

[0017] 音乐处理方法还可以包括以下步骤：通过比较针对音乐数据生成的各个特征模式与任意音乐数据的特征模式来计算针对各个特征模式的相似度；并且基于与多个特征模式之中计算的相似度最高的特征模式对应的编码开始位置来校正音乐数据的偏移或者与音乐数据有关的元数据的偏移。

[0018] 根据本发明的另一实施例，提供了一种包括以下步骤的音乐处理方法：获取音乐数据中在时间轴上的窗口宽度中的平均音量第一次超过全部或者部分音乐数据的平均音量时的窗口位置；并且基于音乐数据的窗口位置来校正音乐数据的偏移或者与音乐数据有关的元数据的偏移。

[0019] 根据本发明的另一实施例，提供了一种音乐处理装置，该装置包括：代码值确定单元，该代码值确定单元根据音乐数据中在时间轴上从编码开始位置开始的每一定时段中的音量转变来确定代码值；以及特征模式生成单元，该特征模式生成单元使用由代码值确定单元确定的在多个时段上的系列代码值来生成表明音乐数据的音量转变特征的特征模式。

[0020] 根据本发明的另一实施例，提供了一种使控制信息处理装置的计算机作为以下单元来工作的程序：代码值确定单元，该代码值确定单元根据音乐数据中在时间轴上从编码开始位置开始的每一定时段中的音量转变来确定代码值；以及特征模式生成单元，该特征模式生成单元使用由代码值确定单元在多个时段上确定的系列代码值来生成表明音乐数据的音量转变特征的特征模式。

[0021] 根据本发明的上述音乐处理方法、音乐处理装置和程序，无论偏移差异如何仍然可以使用音乐数据或者音乐数据的元数据。

附图说明

[0022] 图 1 是示出音乐数据、元数据与再混合数据之间关系的说明图；

[0023] 图 2 是用于说明其中在不同环境中记录音乐的情形的说明图；

- [0024] 图 3 是示出根据第一实施例的音乐处理装置的逻辑配置的框图；
- [0025] 图 4 是用于说明根据第一实施例的参考平均音量计算过程的说明图；
- [0026] 图 5 是用于说明根据第一实施例的编码开始位置确定过程的说明图；
- [0027] 图 6 是用于说明根据第一实施例的代码值确定过程的说明图；
- [0028] 图 7 是用于说明根据第一实施例从特征模式生成单元输出的数据集的说明图；
- [0029] 图 8 是示出根据第一实施例的特征模式生成过程例子的流程图；
- [0030] 图 9 是示出根据第一实施例的编码开始位置确定过程例子的流程图；
- [0031] 图 10 是示出根据第一实施例的代码值确定过程例子的流程图；
- [0032] 图 11 是用于说明根据第一实施例的音乐识别过程的说明图；
- [0033] 图 12 是示出根据第一实施例的音乐识别过程例子的流程图；
- [0034] 图 13 是用于说明根据第一实施例的偏移校正过程的说明图；
- [0035] 图 14 是示出根据第一实施例的偏移校正过程例子的流程图；
- [0036] 图 15 是示出根据第二实施例的音乐处理装置的逻辑配置的框图；
- [0037] 图 16 是用于说明根据第二实施例的编码开始位置确定过程的说明图；
- [0038] 图 17 是用于说明根据第二实施例的代码值确定过程的说明图；
- [0039] 图 18 是用于说明根据第二实施例的从特征模式生成单元输出的数据集的说明图；
- [0040] 图 19 是示出根据第二实施例的特征模式生成过程例子的流程图；
- [0041] 图 20 是用于说明根据第二实施例的音乐识别过程的说明图；
- [0042] 图 21 是示出根据第二实施例的音乐识别过程和偏移校正过程例子的流程图；并且
- [0043] 图 22 是示出通用计算机的配置例子的框图。

具体实施方式

[0044] 下文将参照附图具体地描述本发明的优选实施例。注意在本说明书和附图中用相同标号表示功能和结构基本上相同的结构单元，并且省略对这些结构单元的重复说明。

[0045] 将按以下顺序描述本发明。

[0046] 1. 时序元数据的问题

[0047] 2. 第一实施例

[0048] 3. 第二实施例

[0049] <1. 时序元数据的问题>

[0050] 将参照图 1 描述应用于音乐数据的时序元数据。图 1 是用于说明音乐数据、元数据和再混合数据之间的关系的关系的说明图。

[0051] 在图 1 中沿着各自时间轴以音频信号波形的形式示出音乐 Ga 的音乐数据 Da 和音乐 Gb 的音乐数据 Db。音乐数据 Da 和 Db 可以是任何格式、比如 WAVE 或者 MP3 (MPEG 音频第 3 层) 的音频数据。

[0052] 元数据 Ma 附加于音乐数据 Da，而元数据 Mb 附加于音乐数据 Db。在图 1 中所示的例子中，各个元数据 Ma 和 Mb 包括和弦进展数据（比如“C#”、“C6”、“C7”和“E7”）、由与时间轴垂直的线表明的节拍位置数据以及由偏移 Fa 和 Fb 表明的偏移数据。注意在本说明书

中,偏移代表在音乐数据开始点与音乐演奏开始点(也就是在再现期间检测到声音的第一时间点)之间的时间差。

[0053] 图 1 也示出再混合数据 Rab。再混合数据代表与用于通过组合多段音乐数据的全部或者部分来创作作为整体的新音乐的“配方”(组成)有关的数据。在图 1 中,作为再混合数据 Rab 的内容例子示出如下组成:其中在音乐数据 Da 的再现期间使音乐数据 Da 渐弱而在音乐数据 Da 的渐弱期间使音乐数据 Db 渐强。这里,再混合数据不应限于图 1 中所示的例子,并且可以是诸如播放时间规定、声音音量规定、调制、重复等与音频信号编辑有关的任何数据。

[0054] 这里,为了基于再混合数据 Rab 来准确地再现音乐数据 Ga 和 Gb,希望偏移 Fa 和 Fb 分别准确地表明音乐数据 Da 和 Db 的偏移。另一方面,例如当元数据 Mb 中包括的偏移 Fb 不同于音乐数据 Db 的偏移时,在根据再混合数据 Rab 使音乐数据 Db 渐强时可能没有正确地使音乐数据 Db 渐强。

[0055] 图 2 示出其中在不同环境 E1 和 E2 中记录音乐 Ga 的情况。如图 2 中所示,例如由于数据读取设备的差别或者编码参数的差别,在环境 E1 中记录的音乐 Ga 的音乐数据 Da' 的偏移与在环境 E2 中记录的音乐 Ga 的音乐数据 Da" 的偏移之间有差别。因此,基于音乐数据 Da 预先生成的元数据 Ma 的偏移 Fa 与音乐数据 Da' 的偏移匹配,但是并不与音乐数据 Da" 的偏移匹配。换言之,单独获得的元数据 Ma 不再适用于在环境 E2 中记录的音乐数据 Da"。

[0056] 在其中音乐数据用途已经变得多样化并且元数据很可能被独自分发的当前情形中,由这样的记录环境中的差别等造成的偏移差异是一个不可避免的问题。有鉴于此,希望实现一种在本说明书中描述的、无论此类偏移差异的影响如何仍然能够使用音乐数据或者音乐数据的元数据的设备。

[0057] <2. 第一实施例>

[0058] 图 3 是用于说明根据本发明第一实施例的音乐处理装置 10 的逻辑配置的框图。在图 3 中,音乐处理装置 10 包括存储单元 12、特征模式生成单元 14、开始位置确定单元 16、代码值确定单元 18、音乐识别单元 20 和偏移校正单元 22。

[0059] [存储单元]

[0060] 存储单元 12 被配置成具有存储设备如硬盘和闪存并且存储将要在音乐处理装置 10 中处理的音乐数据、在后文描述的过程中生成的特征模式等。

[0061] [特征模式生成单元]

[0062] 特征模式生成单元 14 从存储单元 12 获取音乐数据并且生成特征模式,该特征模式表明获取的音乐数据的音量转变特征。在本实施例中,将特征模式给定为代码值数组,其中音乐数据中在时间轴上的每一定时段中以先前时段为参照的相对音量转变或者音量改变被编码。当生成特征模式时,特征模式生成单元 14 调用开始位置确定单元 16 并且让开始位置确定单元 16 确定时间轴上将要开始编码的位置、也就是编码开始位置。

[0063] [开始位置确定单元]

[0064] 开始位置确定单元 16 从特征模式生成单元 14 接收音乐数据,并且确定时间轴上将要对音乐数据的音量转变特征开始进行编码的编码开始位置。例如,可以基于时间轴上的窗口宽度之中音量比音乐数据的全部或者部分的平均音量更大的第一窗口位置来确定

编码开始位置。下文将参照图 4 和图 5 描述开始位置确定单元 16 对编码开始位置的确定过程的例子。

[0065] 如图 4 中所示,开始位置确定单元 16 计算音乐数据的预定范围的、用作确定编码开始位置的参考的参考平均音量 V_m 。在图 4 中将用于计算参考平均音量 V_m 的预定范围表示为从 T_a 到 T_b 的范围。从 T_a 到 T_b 的范围可以包括全部音乐数据或者部分音乐数据。可以用正或者负代码代表 音乐数据的量化音量;然而,参考平均音量 V_m 可以是音量的绝对值之中的平均值或者音量的平方值之中的平均值。

[0066] 如图 5 中所示,开始位置确定单元 16 从用来确定参考平均音量 V_m 的范围的头部位置 T_a 开始并且沿着时间轴搜索在时间轴上具有窗口宽度 W_s 的窗口的平均音量 V_w 超过参考平均音量 V_m 的第一位置。

[0067] 例如,在图 5-5A 中所示窗口位置,窗口宽度 W_s 的平均音量 V_w 小于参考平均音量 V_m 。开始位置确定单元 16 因此在正方向上在时间轴上移动下一窗口位置并且再次比较下一窗口宽度 W_s 的平均音量 V_w 与参考平均音量 V_m 。希望以更小单位来移动窗口位置,以便改进编码开始位置确定的可重复性。例如,可以与音乐数据采样的各样本对应地移动窗口位置。

[0068] 在上述过程之后,例如假设在正方向上沿着时间轴移动窗口位置,并且窗口宽度 W_s 的平均音量 V_w 在窗口的开始点到达图 5-5B 中所示位置 T_s 时第一次变得大于参考平均音量 V_m 。开始位置确定单元 16 检测到位置 T_s 并且将位置 T_s 作为编码开始位置输出到特征模式生成单元 14。

[0069] 当确定编码开始位置 T_s 时,特征模式生成单元 14 调用代码值确定单元 18,以让代码值确定单元 18 确定代码值从而针对从编码开始位置 T_s 起的各个一定时段生成特征模式。在本实施例中,用于生成特征模式的代码值被假设为与每一定时段中以先前时段为参照的音量转变对应的值。

[0070] [代码值确定单元]

[0071] 图 6 是用于说明代码值确定单元 18 的代码值确定过程的例子的说明图。图 6 示出从编码开始位置 T_s 开始的具有时段宽度 W_e 的九个编码时段。另外,以代表相应编码时段的矩形形状的高度形式示出相应时段的在各编码时段中计算的平均音量。在本实施例中,代码值确定单元 18 依次地计算图 6 中所示相应时段的平均音量,并且基于相应时段的计算的平均音量是大于还是小于相应前一时段的平均音量来确定代码值。

[0072] 在图 6 中所示例子中,在对象时段中的平均音量大于先前时段的平均音量时,代码值确定为“1”,而在对象时段中的平均音量等于或者小于先前时段的平均音量时,代码值确定为“0”。例如,由于第二时段的平均音量大于在编码开始位置 T_s 开始的第一时段的平均音量,所以第一代值(最左边的代码值)是“1”。类似地,由于第三时段的平均音量大于第二时段的平均音量,所以第二代值也是“1”。另一方面,由于第四时段的平均音量小于第三时段的平均音量,所以第三代值是“0”。

[0073] 特征模式生成单元 14 组合相应时段的由代码值确定单元 18 确定的代码值,并且将特征模式生成为一系列代码值。在图 6 的例子中,基于相应编码时段的确定的代码值来生成特征模式 CP“11011010”。在这一例子中,为了便于描述,示出九个编码时段;然而,编码时段的数目(也就是特征模式中的位数)优选地设置成足以使用特征模式相对于彼此识

别音乐数据的数目。

[0074] 为了生成特征模式,只要代码值代表沿着时间轴的相对音量转变,就不仅可以使使用二元值也可以使用多元值(multiple values)。另外,作为如图6中所示的与对象时段和前一时段之间的音量转变对应的值的取代值,代码值可以是与音量转变特征(比如在对象时段的平均音量与参考平均音量之间的音量比较)对应的任何值。

[0075] 在上述过程中生成的特征模式 CP 以及用于生成特征模式 CP 的窗口宽度 W_s 和编码时段宽度 W_e 将如下所述用来识别音乐数据。另外,编码开始位置 T_s 用来校正偏移。如图7中所示,特征模式生成单元 14 将特征模式 CP、窗口宽度 W_s 、编码时段宽度 W_e 和编码开始位置 T_s 作为数据集来对待,并且在使用例如用于标识音乐数据的音乐标识符将数据集与音乐数据关联之后将数据集存储到存储单元 12。

[0076] [对过程流程的说明:特征模式的生成]

[0077] 图8至图10是用于说明根据本实施例的音乐处理装置10生成特征模式的流程的流程图。

[0078] 如图8中所示,特征模式生成单元 114 首先从存储单元 12 获取音乐数据 (S110)。特征模式生成单元 14 调用开始位置确定单元 16,并且开始位置确定单元 16 计算参考区域 T_a 到 T_b 中的参考平均音量 V_m (S120)。然后,开始位置确定单元 16 基于参考平均音量 V_m 和窗口宽度 W_s 来确定编码开始位置 T_s (S130)。

[0079] 图9示出图8中的确定编码开始位置这一步骤 S130 的具体流程。

[0080] 如图9中所示,首先基于参考区域的开始点 T_a 来确定窗口的初始位置 (S302)。计算在该窗口位置的窗口的平均音量 V_w (S304)。然后比较窗口的所计算的平均音量 V_w 与参考平均音量 V_m (S306)。这里,例如当窗口的平均音量 V_w 等于或者小于参考平均音量 V_m 时,例如将窗口位置移动到下一样本,并且该过程返回到 S304 (S308)。另一方面,在步骤 S306 中窗口的平均音量 V_w 大于参考平均音量 V_m 时,基于这时的窗口位置来确定编码开始位置 T_s (S310)。

[0081] 回到图8,生成特征模式的流程将继续。当确定编码开始位置 T_s 时,特征模式生成单元 14 基于编码开始位置 T_s 和编码时段宽度 W_e 来生成特征模式 CP (S140)。

[0082] 图10示出图8中的生成特征模式这一步骤 S140 的具体流程。

[0083] 如图10中所示,首先将第一时段的平均音量设置为先前时段的平均音量 V_p 的初始值 (S402)。接着计算第“n”个时段的平均音量 V_n (S404)。在第一次计算中,“n”是“2”。然后比较第“n”个时段的所计算的平均音量 V_n 与先前时段的平均音量 V_p (S406)。这里,例如当先前时段的平均音量 V_p 等于或者大于第“n”个时段的平均音量 V_n 时,将第“n”个时段的代码值 C_n 确定为“0” (S408)。当第“n”个时段的平均音量 V_n 更大时,将第“n”个时段的代码值 C_n 确定为“1” (S410)。此后确定是否完成对所有时段的代码值的确定 (S412)。当没有完成对所有时段的代码值的确定时,将第“n”个时段的平均音量赋值为先前时段的平均音量 V_p ,用“n+1”取代“n”,并且该过程返回到 S404 (S414)。另一方面,当在步骤 S412 中完成对所有时段的代码值的确定时,基于相应时段的代码值 C_n ($n = 2, 3, \dots$) 来生成特征模式 CP (S416)。

[0084] 再次参照图8,当生成特征模式 CP 时,特征模式生成单元 14 将特征模式 CP、窗口宽度 W_s 、编码时段宽度 W_e 和编码开始位置 T_s 作为一个数据集存储到存储单元 12 (S150)。

[0085] 参照图 4 至图 10, 已经主要描述根据本实施例的特征模式生成过程。在本实施例中生成的特征模式如上所述是音乐数据在时间轴上的被编码的相对音量转变的代码值数组。音乐数据的音量可能由于比如数据读取设备的差别或者编码参数的差别这样的环境差别而变化。然而, 这些环境差别大体上均匀地影响全部音乐数据。因此, 当基于时间轴上的相对音量转变来确定代码值时, 消除了这些环境差别的影响。有鉴于此, 可以理解在本实施例中生成的特征模式无论环境差别的影响如何都适合于识别在不同环境中记录的多段相同音乐。

[0086] 另外在本实施例中, 也基于沿着时间轴的相对音量转变来确定用于生成特征模式的编码开始位置。换言之, 在使用相同音乐数据的情况下, 编码开始位置的差别实际上仅由音乐数据在时间轴上的演奏位置、也就是 偏移所造成。利用这一配置, 可以理解当使用相同音乐数据或者该音乐数据的元数据时, 在本实施例中确定的编码开始位置可以用于校正偏移差异。

[0087] [音乐识别单元]

[0088] 接着回到图 3, 将描述配置成基于上述特征模式来识别音乐的音乐识别单元 20。

[0089] 音乐识别单元 20 比较针对某些音乐数据生成的特征模式与任意音乐数据的特征模式, 以计算这些段音乐数据的特征模式的相似度。音乐识别单元 20 将要比较的特征模式可以是由相同音乐处理装置 10 生成的特征模式或者由不同设备生成的特征模式。然而, 假设将相同窗口宽度 W_s 和编码时段宽度 W_e 用于生成特征模式。

[0090] 例如, 假设已知音乐数据 D 的特征模式 CP 以及用来生成特征模式 CP 的窗口宽度 W_s 和编码时段宽度 W_e 。在这一情况下, 为了确定另一音乐数据 D_a 与音乐数据 D 的同一性, 基于窗口宽度 W_s 和编码时段宽度 W_e 来生成音乐数据 D_a 的特征模式 CP_a , 并且比较生成的特征模式 CP_a 与特征模式 CP。

[0091] 图 11 是用于说明由音乐识别单元 20 执行的音乐识别过程的例子的说明图。

[0092] 在图 11 中, 特征模式生成单元 14 使用相同窗口宽度 W_s 和编码时段宽度 W_e 来事先生成音乐数据 D 的特征模式 CP、音乐数据 D_a 的特征模式 CP_a 和音乐数据 D_b 的特征模式 CP_b 。音乐识别单元 20 例如从存储单元 12 读取这些特征模式。

[0093] 然后, 为了确定在音乐数据 D 与音乐数据 D_a 之间的同一性, 音乐识别单元 20 计算特征模式 CP 与特征模式 CP_a 的相应位的异或 (EXOR)。如图 11 中所示, 特征模式 CP 与特征模式 CP_a 的异或是“00001100”。另外, 为了确定在音乐数据 D 与音乐数据 D_b 之间的同一性, 音乐识别单元 20 计算特征模式 CP 与特征模式 CP_b 的相应位的异或。特征模式 CP 与特征模式 CP_b 的异或是“01101101”。

[0094] 这里, 计算的异或中的位“0”表明特征模式的位的代码值相同。位“1”表明特征模式的位的代码值不同。换言之, 计算的异或中的位“0”的比例越高, 表明特征模式之间的相似度越高。

[0095] 因此, 音乐识别单元 20 可以例如使用两个特征模式之间的计算的异 或中的位“0”的数目与总位数之比作为特征模式的相似度。在图 11 的例子中, 基于特征模式 CP 与特征模式 CP_a 的计算的异或“00001100”将这些特征模式之间的相似度 SIM_a 计算为 $6[\text{位}]/8[\text{位}] = 75\%$ 。类似地, 基于特征模式 CP 与特征模式 CP_b 的计算的异或“01101101”将这些特征模式之间的相似度 SIM_b 计算为 $3[\text{位}]/8[\text{位}] = 37.5\%$ 。

[0096] 另外,虽然在图 11 中未示出,但是可以基于二项式分布中表明在两个特征模式之间代码值匹配的位“0”和表明代码值不匹配的位“1”的出现概率,而不是表明匹配的代码值的位“0”的简单比例,来计算相似度。更具体而言,例如当表明匹配的代码值的位“0”的简单比例假设为 P_0 时,根据其中试验次数 = 总位数并且成功概率 = 50% 的二项式分布中的概率 X_0 (其使得随机变量 $X \geq P_0$) 来给出基于二项式分布的相似度 (注意在这一情况下,概率 X_0 越低,表明相似度越高)。

[0097] 另外,在计算的异或中表明代码值匹配的位“0”的连续部分的最大长度可以用来计算特征模式的相似度。作为例子,考虑一种对在两个不同环境中记录的相同音乐的特征模式进行比较的情况。当在记录一段音乐数据期间仅在一部分中有噪声时,该噪声部分的特征模式的代码值不匹配的概率为 50%。在无噪声的其他部分中,理想地在两个特征模式之间的所有其他代码值匹配。另一方面,当比较不同音乐的特征模式时,不仅在特定范围如噪声部分中而是在整个特征模式中发现不匹配的代码值。因此,例如当表明代码值匹配的连续位“0”的最大长度用作特征模式的相似度时,就噪声出现而言可以更稳健地执行对音乐同一性的确定。

[0098] 音乐识别单元 20 可以例如将通过任何上述方法计算的特征模式的相似度与预定阈值比较,以确定两段音乐数据的同一性。另外,音乐识别单元 20 可以使用连接到音乐处理装置 10 的显示设备向用户通知计算的特征模式的相似度,以让用户确定音乐的同一性。

[0099] [过程流程的说明 : 音乐识别]

[0100] 图 12 是用于说明根据本实施例的音乐处理装置 10 的音乐识别流程的流程图。

[0101] 如图 12 中所示,音乐识别单元 20 首先从存储单元 12 获取两个特征模式 CP_x 和 CP_y (S610)。音乐识别单元 20 然后计算特征模式 CP_x 和 CP_y 的异或 (S620)。然后,音乐识别单元 20 基于计算的异或来计算特征模式 CP_x 和 CP_y 的相似度 SIM (S630)。此后,音乐识别单元 20 比较 计算的相似度 SIM 与预定阈值,以例如确定音乐的同一性,或者显示相似度 SIM 以让用户确定音乐的同一性 (S640)。

[0102] 上文已经参照图 11 和图 12 描述了根据本实施例的音乐处理装置 10 的音乐识别过程。如上所述,通过使用表明音量转变特征的特征模式来执行根据本实施例的音乐识别过程。因此,无论例如由环境差别造成的偏移差异等的影响如何,仍然能够以实用的准确性来执行音乐识别。

[0103] [偏移校正单元]

[0104] 将描述一种基于上述音乐识别过程的结果和用于生成特征模式的编码开始位置来校正从数据开始点到演奏开始点的偏移差异的过程。

[0105] 偏移校正单元 22 基于与具有高相似度的音乐数据的特征模式对应的上述编码开始位置来校正音乐数据或者元数据的偏移。图 13 是用于说明偏移校正单元 22 的偏移校正过程的例子的说明图。

[0106] 图 13-13A 示出了音乐 G_a 的音乐数据 D_a 和应用于音乐数据 D_a 的元数据 M_a 的波形。图 13-13A 也示出了元数据 M_a 中包括的音乐数据 D_a 的偏移 T_o 与音乐数据 G_a 的编码开始位置 T_s 。

[0107] 另外,图 13-13B 示出音乐 G_a 的在与音乐数据 D_a 的记录环境不同的环境中记录的音乐数据 D_a' 的波形。参照音乐数据 D_a' 的波形,音乐数据 D_a' 的演奏开始定时晚于音乐数

据 Da 的演奏开始定时。换言之,由于环境差别,在音乐数据 Da 与音乐数据 Da' 的偏移之间存在间隙。因此,当音乐数据 Da 的元数据 Ma 简单地用于音乐数据 Da' 时,元数据 Ma 中包括的比如节拍位置以及和弦进展这样的时序数据并不对应于音乐数据 Da' 的时序数据。

[0108] 偏移校正单元 22 获取由开始位置确定单元 16 确定的针对音乐数据 Da' 的编码开始位置 Ts', 并且根据与音乐数据 Ga 的编码开始位置 Ts 的差别在元数据 Ma 应用于音乐数据 Da' 时校正偏移。使用以下方程来计算校正的偏移 To'。

[0109] [方程 1]

$$[0110] \quad To' = To + (Ts' - Ts) \quad (1)$$

[0111] 如图 13 中所示,由于元数据 Ma 的偏移在应用于音乐数据 Da' 之前被校正,所以元数据 Ma 中包括的比如节拍位置以及和弦进展这样的时序数据可以与音乐数据 Da' 的时序数据匹配。

[0112] 参照图 13,已经描述了使用校正值 Ts' - Ts 来校正元数据 Ma 的偏移的例子。然而,可以代之以使用校正值 Ts' - Ts 来校正音乐数据 Da' 的偏移。换言之,当再现音乐数据 Da' 时,可以使播放时间的开始点跳跃达校正值 Ts' - Ts 的长度,以与简单再现的元数据 Ma 的演奏匹配。

[0113] 另外,已经基于预先已知根据相同音乐 Ga 生成音乐数据 Da 和音乐数据 Da' 这一假设来进行上述描述。然而,例如可以在多段音乐数据选项之中自动选择或者由用户人工选择与对象音乐数据的特征模式相似度高的音乐数据,并且可以基于所选音乐数据的编码开始位置来校正偏移。

[0114] [过程流程的说明 : 偏移校正]

[0115] 图 14 是用于说明根据本实施例的音乐处理装置 10 的偏移校正流程的流程图。

[0116] 如图 14 中所示,偏移校正单元 22 首先获取对象音乐的编码开始位置 Ts' (S710)。偏移校正单元 22 获取与对象音乐的特征模式相比具有特征模式的高相似度的音乐数据的编码开始位置 Ts (S720)。编码开始位置 Ts 和 Ts' 例如由开始位置确定单元 16 预先确定。另外,偏移校正单元 22 计算偏移的校正值 Ts' - Ts (S730)。然后,偏移校正单元 22 基于校正值 Ts' - Ts 来校正该校正对象音乐数据的偏移或者校正对象音乐数据的元数据的偏移 (S740)。

[0117] 参照图 13 和图 14,上文已经描述根据本实施例的音乐处理装置 10 的偏移校正过程。基于实质上表明一段音乐数据在时间轴上的演奏位置的编码开始位置来执行根据本实施例的偏移校正过程。利用这一配置,可以在校正由环境差别造成的偏移差异之后使用音乐数据或者音乐数据的元数据。

[0118] 参照图 3 至图 14 具体描述了本发明的第一实施例。在本发明的第一实施例中,针对一段音乐数据生成表明时间轴上的相对音量转变特征的单个特征模式。这里,如图 6 的说明中所见,基于由开始位置确定单元 16 确定的编码开始位置来确定特征模式中包括的代码值。因此,有可能针对一段音乐数据生成与多个编码开始位置对应的多于一个的特征模式,以改进音乐同一性确定的质量。有鉴于此,下文将描述如下例子作为本发明的第二实施例,在该例子中针对一段音乐数据基于使用不同窗口位置确定的多个编码开始位置来生成多个特征模式。

[0119] <3. 第二实施例 >

[0120] 图 15 是用于说明根据本发明的第二实施例的音乐处理装置 30 的逻辑配置的框图。如图 15 中所示,音乐处理装置 30 包括存储单元 12、特征模式生成单元 34、开始位置确定单元 36、代码值确定单元 38、音乐识别单元 40 和偏移校正单元 22。

[0121] [特征模式生成单元]

[0122] 根据本实施例,特征模式生成单元 34 从存储单元 12 获取音乐数据并且生成表明所得音乐数据的音量转变特征的多个特征模式。当生成特征模式时,特征模式生成单元 34 首先调用开始位置确定单元 36 并且让开始位置确定单元 36 使用不同窗口宽度来确定多个编码开始位置。

[0123] [开始位置确定单元]

[0124] 图 16 是用于说明根据本实施例的开始位置确定单元 36 的编码开始位置确定过程的说明图。

[0125] 如图 16 中所示,开始位置确定单元 36 计算参考平均音量 V_m ,然后确定在时间轴上具有窗口宽度 W_{s_1} 的窗口的平均音量 V_{w_1} 第一次超过参考平均音量 V_m 时的位置 T_{s_1} (16A)。接着,开始位置确定单元 36 确定在时间轴上具有窗口宽度 W_{s_2} 的窗口的平均音量 V_{w_2} 第一次超过参考平均音量 V_m 时的位置 T_{s_2} (16A),其中该窗口宽度 W_{s_2} 不同于窗口宽度 W_{s_1} 。如上所述,在本实施例中,开始位置确定单元 36 基于不同窗口宽度 W_{s_i} 来确定多个编码开始位置 T_{s_i} ($i = 1, 2, \dots$),并且将确定的多个编码开始位置 T_{s_i} 输出到特征模式生成单元 34。

[0126] [代码值确定单元]

[0127] 图 17 是用于说明根据本实施例的代码值确定单元 38 的代码值确定过程的说明图。

[0128] 图 17-17A 示出在编码开始位置 T_{s_1} 开始的具有时段宽度 W_e 的九个编码时段。图 17-17B 也示出在编码开始位置 T_{s_2} 开始的具有时段宽度 W_e 的九个编码时段。关于各个编码时段,编码值确定单元 38 根据各个时段的平均音量是大于还是小于各自的前一时段来确定代码值。

[0129] 特征模式生成单元 34 组合一系列编码时段内的各个时段的确定代码值,并且生成各个编码开始位置 T_{s_i} 的特征模式 CP_i 。在图 17 的例子中,针对编码开始位置 T_{s_1} 生成特征模式 CP_1 “11011010”,而针对编码开始位置 T_{s_2} 生成特征模式 CP_2 “00111010”。

[0130] 在上述过程中生成的特征模式 CP_i 以及用于生成特征模式 CP_i 的窗口宽度 W_{s_i} 和编码时段宽度 W_e 用于后文描述的音乐数据识别。另外,编码开始位置 T_{s_i} 用于偏移校正。如图 18 中的 $i \leq 2$ 这一情况中所示,特征模式生成单元 34 将特征模式 CP_i 、窗口宽度 W_{s_i} 、编码时段宽度 W_e 和编码开始位置 T_{s_i} 作为一个数据集存储到存储单元 12 中。

[0131] [过程流程的说明:多个特征模式的生成]

[0132] 图 19 是用于说明根据本实施例的音乐处理装置 30 的特征模式生成流程的流程图。

[0133] 如图 19 中所示,特征模式生成单元 34 首先从存储单元 12 获取音乐数据 (S1110)。特征模式生成单元 34 调用开始位置确定单元 36,并且特征模式生成单元 34 计算从 T_a 到 T_b 的参考区域内的参考平均音量 V_m (S1120)。然后,开始位置确定单元 36 使用参考平均音量 V_m 和窗口宽度 W_{s_i} ($i = 1, 2, \dots$) 来确定多个编码开始位置 T_{s_i} (S1130)。然后,

特征模式生成单元 34 使用多个编码开始位置 Ts_i 和编码时段宽度 We 来生成多个特征模式 CP_i (S1140)。重复步骤 S1130 和 S1140, 直至生成针对所有窗口宽度 Ws_i 的特征模式 CP_i 。然后, 在生成针对所有窗口宽度 Ws_i 的特征模式 CP_i 之后, 特征模式生成单元 34 将特征模式 CP_i 、窗口宽度 Ws_i 、编码时段宽度 We 和编码开始位置 Ts_i 作为一个数据集存储到存储单元 12 (S1150)。

[0134] [音乐识别单元]

[0135] 将描述配置成基于多个特征模式来识别音乐的音乐识别单元 40。音乐识别单元 40 比较针对音乐数据生成的各个特征模式与针对任意音乐生成的特征模式, 以计算各个特征模式的相似度。这里, 假设比较基于相同窗口宽度 Ws_i 生成的特征模式。

[0136] 图 20 是用于说明音乐识别单元的音乐识别过程的例子的说明图。

[0137] 图 20 示出了针对音乐数据 D 生成的两个特征模式 CP_1 和 CP_2 以及针对音乐数据 $Dref$ 生成的两个特征模式 $CPref_1$ 和 $CPref_2$ 。在这些特征模式之中, 基于相同窗口宽度来分别生成特征模式 CP_1 和 $CPref_1$ 以及特征模式 CP_2 和 $CPref_2$ 。

[0138] 音乐识别单元 40 计算这些特征模式的相应位的异或 (EXOR)。在图 20 的例子中, 特征模式 CP_1 和 $CPref_1$ 的异或是“00000001”。特征模式 CP_2 和 $CPref_2$ 的异或是“11100001”。

[0139] 然后, 音乐识别单元 40 基于计算的异或针对相应窗口宽度计算特征模式的相似度 SIM_1 、 SIM_2 。在图 20 的例子中, 特征模式中的位“0”的简单比例被定义为相似度。在特征模式 CP_1 与 $CPref_1$ 之间的相似度 SIM_1 是 87.5%, 而在特征模式 CP_2 与 $CPref_2$ 之间的相似度 SIM_2 是 50%。

[0140] 如图 20 中所见, 甚至在均根据相同的音乐数据生成的音乐数据 D 与音乐数据 $Dref$ 之间, 特征模式的相似度也可以根据用来确定编码开始位置的窗口宽度而变化。这里, 在多个相似度之中的最高相似度可以用作这两段音乐数据之间的相似度。在这一情况下, 根据图 20 的例子, 由于相似度 SIM_1 高于相似度 SIM_2 , 所以相似度 SIM_1 用作音乐数据 D 与音乐数据 $Dref$ 之间的相似度。

[0141] 用来计算所用的相似度 SIM_1 的特征模式 CP_1 与 $CPref_1$ 的编码开始位置之间的差别与音乐数据 D 和音乐数据 $Dref$ 的偏移差别相关。特征模式 CP_1 和 $CPref_1$ 的编码开始位置的差别因此可以在后文描述的偏移校正过程中用作校正。

[0142] [过程流程的说明 : 音乐识别和偏移校正]

[0143] 图 21 是用于说明根据本实施例的音乐处理装置 30 的音乐识别和偏移校正流程的流程图。

[0144] 如图 21 中所示, 音乐识别单元 40 首先获取待校正的对象音乐数据的特征模式 CP_i ($i = 1, 2, \dots$) (S1210)。音乐识别单元 40 然后获取参考音乐数据的特征模式 $CPref_i$ ($i = 1, 2, \dots$) (S1220)。

[0145] 音乐识别单元 40 针对相应窗口宽度计算特征模式 CP_i 与特征模式 $CPref_i$ 的异或 (S1230)。另外, 基于计算的异或来计算相应窗口宽度的相似度 SIM_i (S1240)。重复步骤 S1230 和 S1240, 直至针对所有窗口宽度计算了相似度 SIM_i 。当完成对所有相似度 SIM_i 的计算时, 音乐识别单元 40 确定“ i ”, 从而获得最高相似度 SIM_i (S1250)。在图 21 中, 假设当 $i = k$ 时获得最高相似度 SIM_i 。

[0146] 偏移校正单元 22 基于与相似度 SIM_k 对应的特征模式 CP_k 的编码开始位置 Ts_k 和

特征模式 $CPref_k$ 的编码开始位置 $Tsref_k$ 来计算偏移校正值 (S1260)。偏移校正单元 2 基于计算的校正值来校正待校正的对象音乐数据的偏移或者音乐数据的元数据 (S1270)。

[0147] 参照图 15 至图 21 描述了本发明的第二实施例。根据本实施例,基于以不同窗口宽度为基础确定的多个编码开始位置针对一段音乐数据生成多个特征模式。然后,基于与生成的多个特征模式之中与参考特征模式的相似度最高的特征模式对应的编码开始位置来校正音乐数据的偏移或者音乐数据的元数据的偏移。这可以改进基于相似度的音乐同一性确定的质量并且提高偏移校正的准确性。

[0148] 根据本说明书中所述的第一和第二实施例的系列过程一般由软件实现。当软件执行系列或者部分过程时,在如图 22 中所示的通用计算机等中执行构成软件的程序。

[0149] 在图 22 中,CPU(中央处理单元)902 控制通用计算机的整体操作。ROM(只读存储器)904 存储部分或者所有所述系列过程的程序或者数据。RAM(随机存取存储器)906 暂时存储 CPU 902 在过程执行期间使用的程序或者数据。

[0150] CPU 902、ROM 904 和 906 RAM 经由总线 910 相互连接。总线 910 还连接到输入/输出接口 912。

[0151] 输入/输出接口 912 是用以将 CPU 902、ROM 904 和 RAM 906 连接到输入设备 920、输出设备 922、存储设备 924、通信设备 926 和驱动器 930 的接口。

[0152] 输入设备 920 例如被配置成经由输入设备如鼠标、键盘或者触板从用户接收指令或者输入信息。输出设备 922 例如被配置成经由显示设备如 CRT(阴极射线管)、液晶显示器和 OLED(有机发光二极管)或者音频输出设备如扬声器向用户输出信息。

[0153] 存储设备 924 由硬盘驱动器、闪存等组成,并且存储程序或者程序数据。通信设备 926 经由网络如 LAN 或者因特网执行通信过程。驱动器 930 根据需要提供给通用计算机并且例如附接有可拆卸介质 932。

[0154] 当根据第一和第二实施例的系列过程由软件执行时,例如图 22 中所示的 ROM 904、存储设备 924 或者可拆卸介质 932 中存储的程序在执行期间由 RAM 906 读取并且由 CPU 902 执行。

[0155] 已经参照附图描述本发明的优选实施例;然而,应当理解本发明不限于那些实施例。本领域技术人员清楚在权利要求的范围内的各种改变和修改是本发明技术范围的一部分。

[0156] 例如,关于本说明书的音乐数据,没有描述立体声数据与单声道数据之间的差别;然而,可以针对各立体声数据单独生成特征模式。另外,可以在组合一对立体声数据之后生成特征模式。

[0157] 此外,在执行第一和第二实施例的各个流程图中所示过程时,可以改变过程的顺序。各个过程步骤可以包括并行或者独立执行的过程。

[0158] 本发明包含与 2008 年 9 月 3 日向日本专利局提交的日本专利申请 JP2008-226344 有关的主题内容,其全部内容通过引用结合于此。

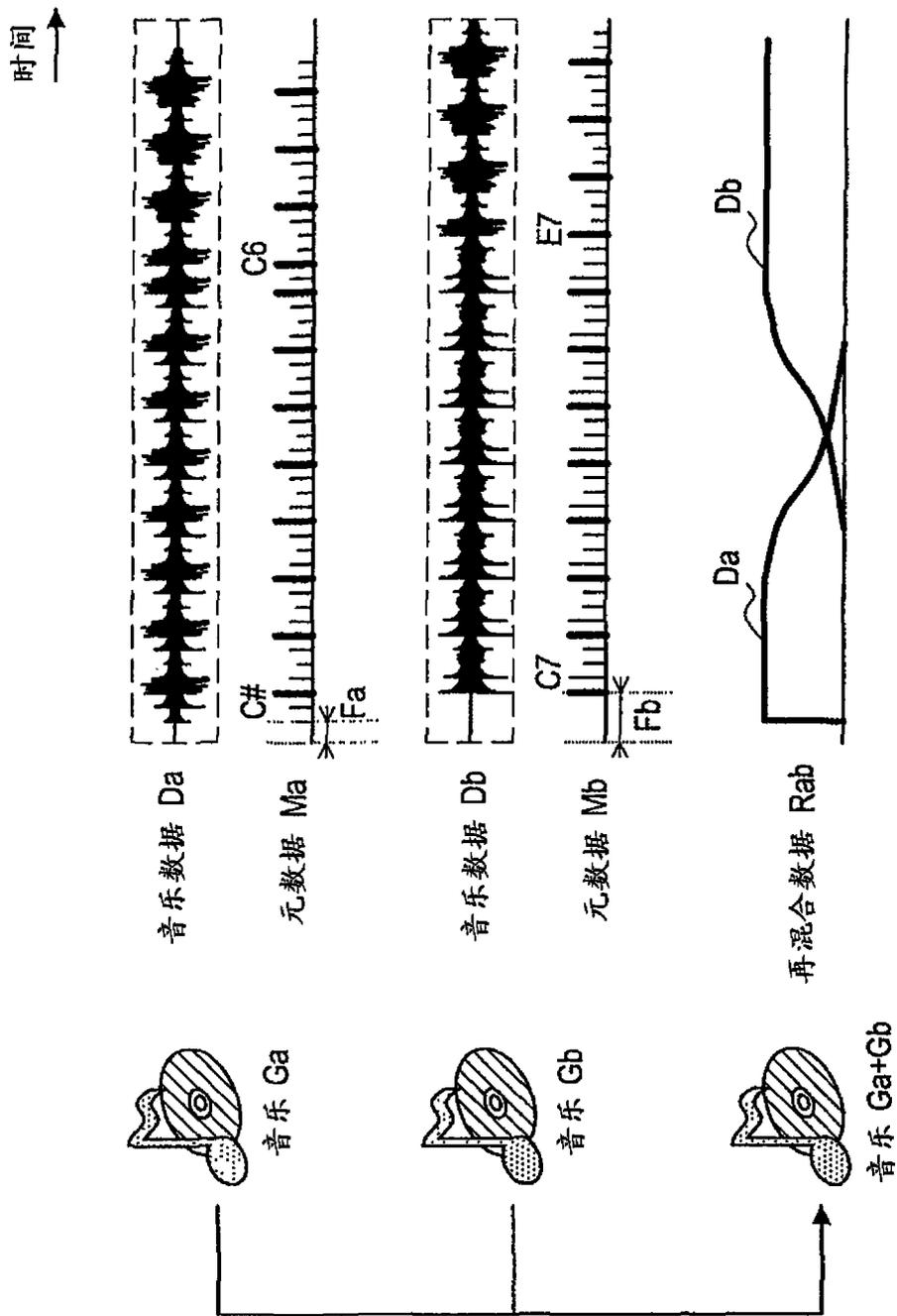


图 1

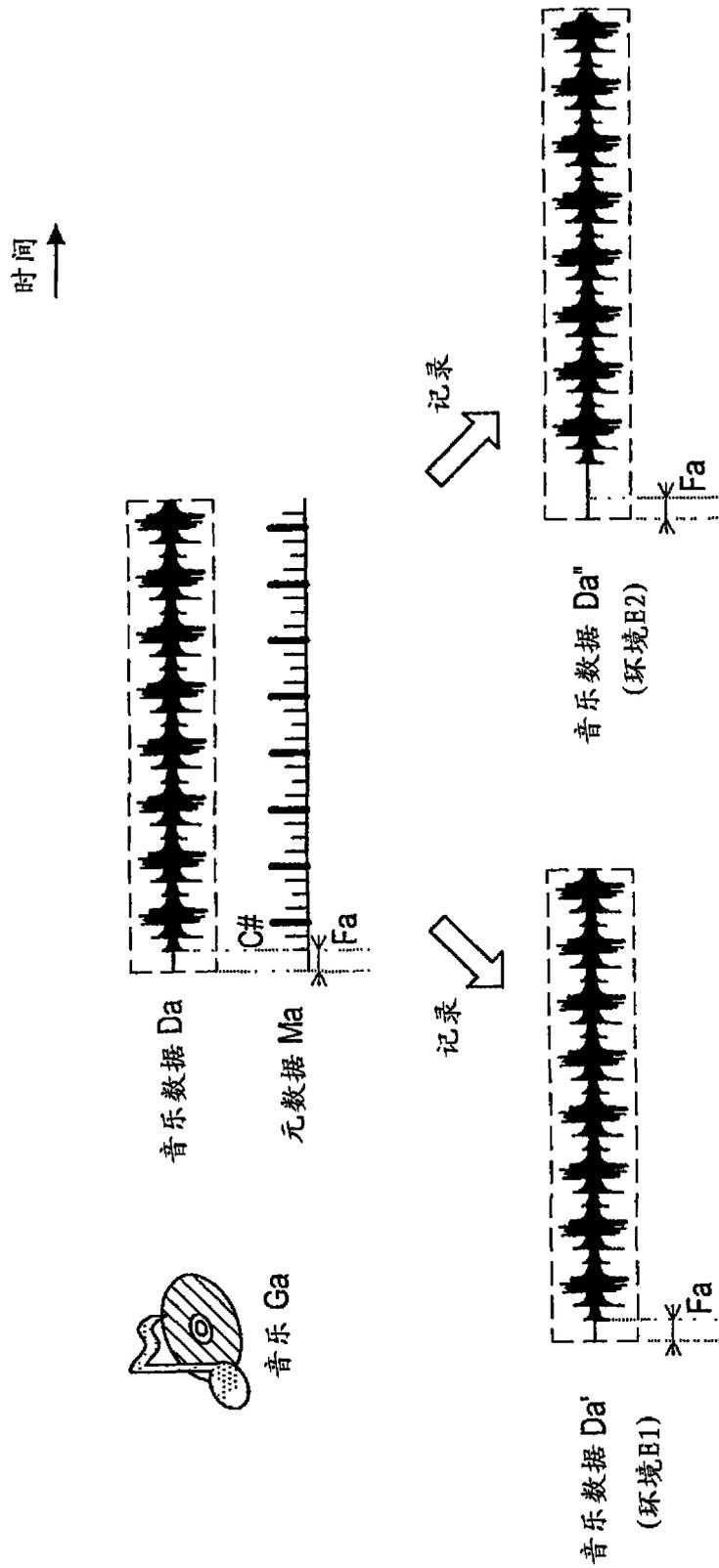


图 2

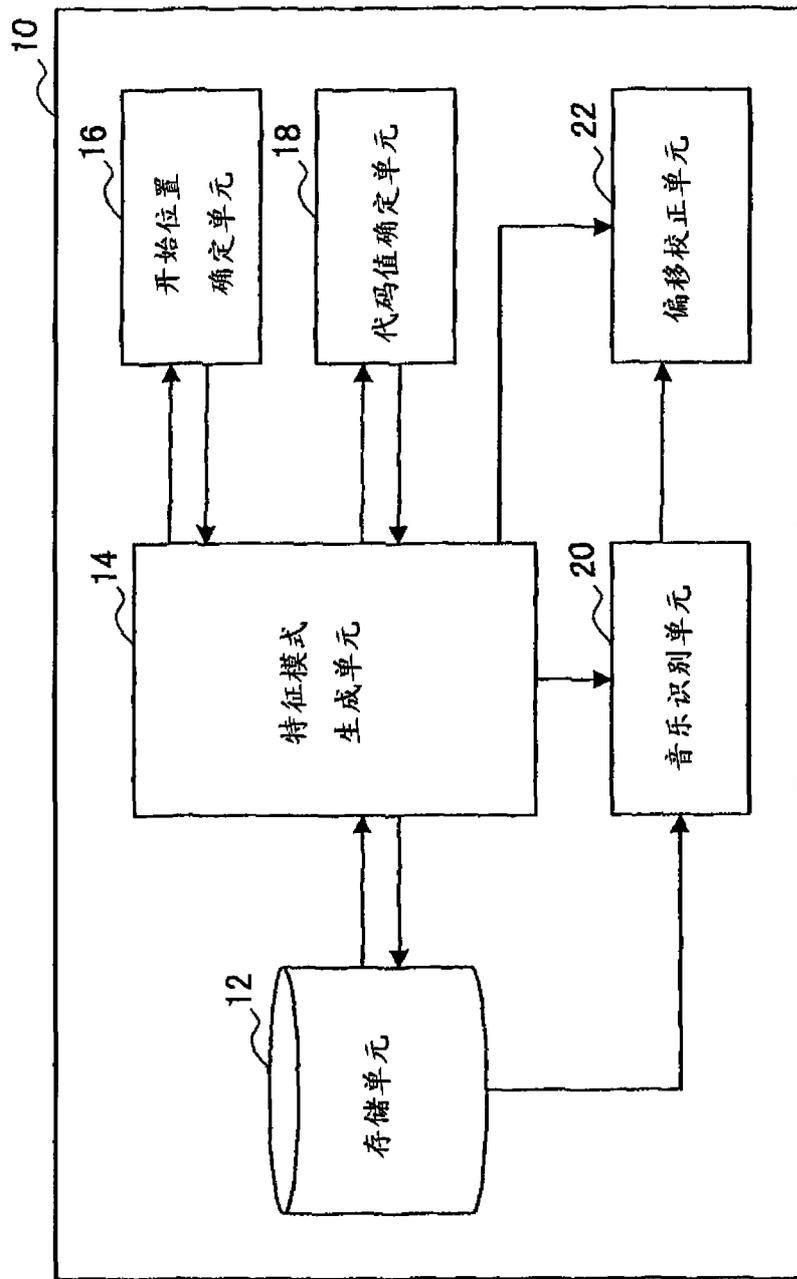


图 3

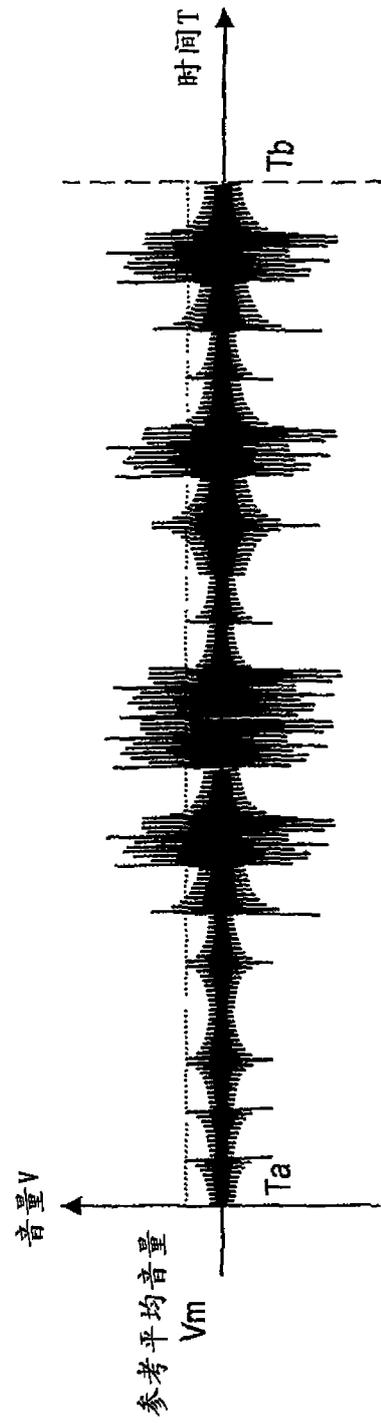


图 4

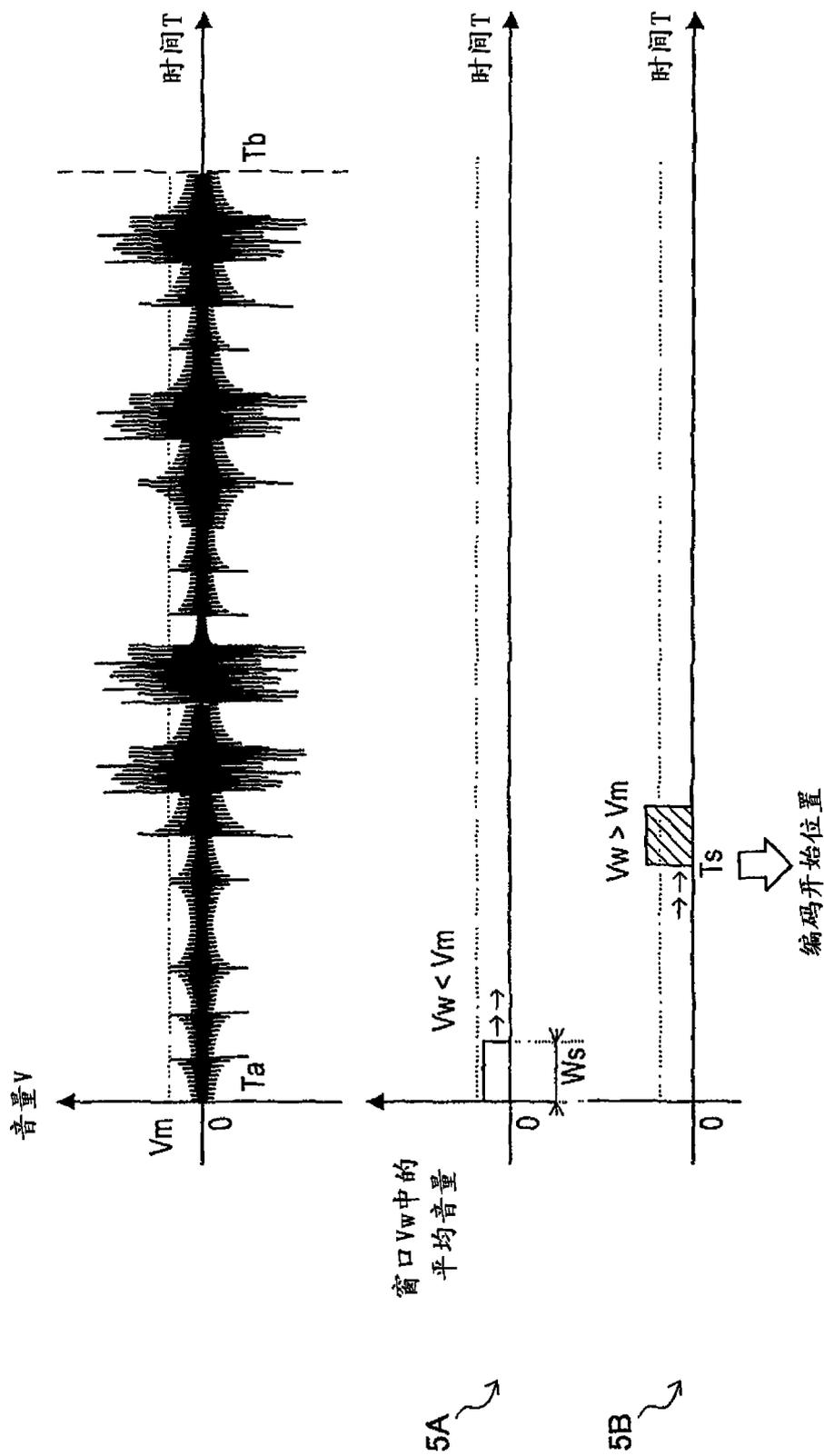


图 5

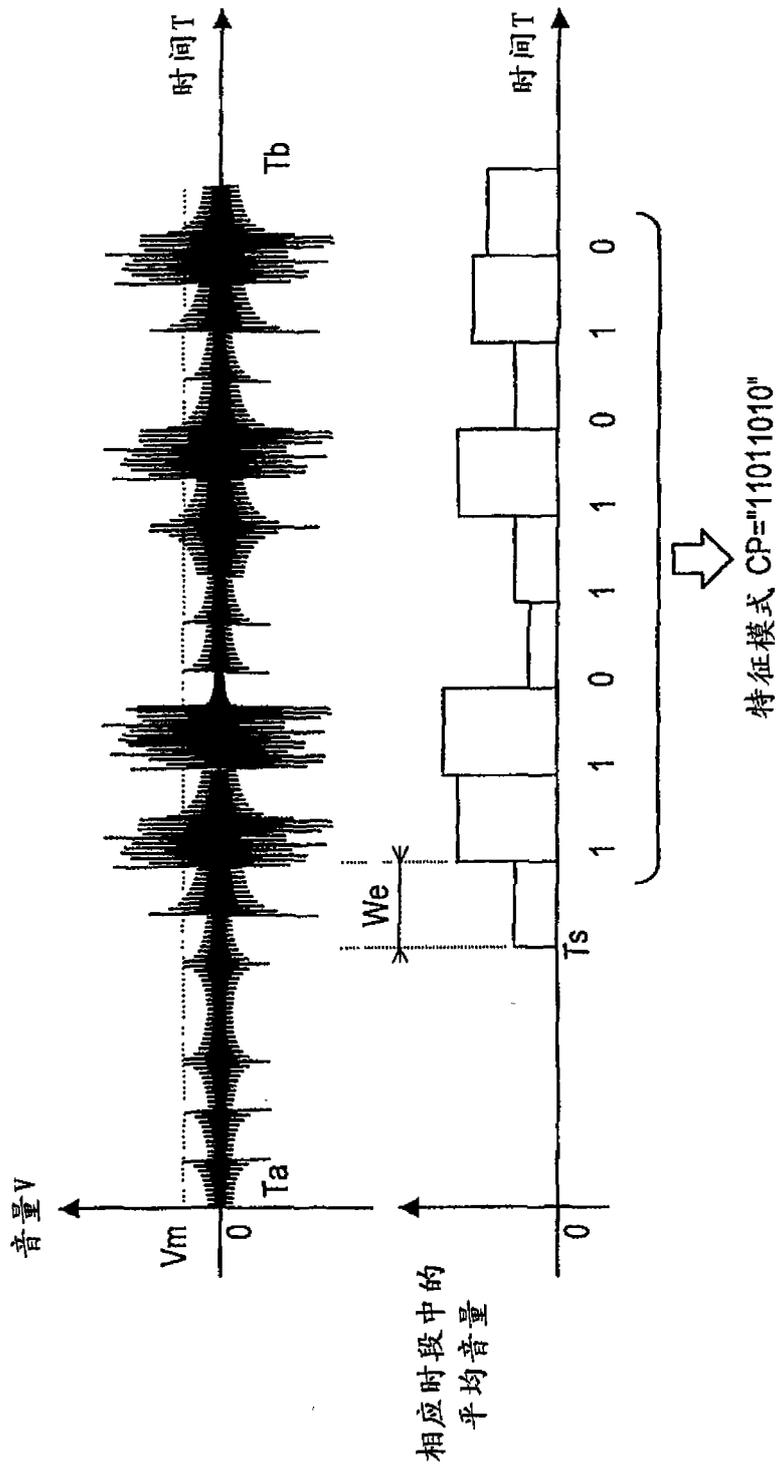


图 6

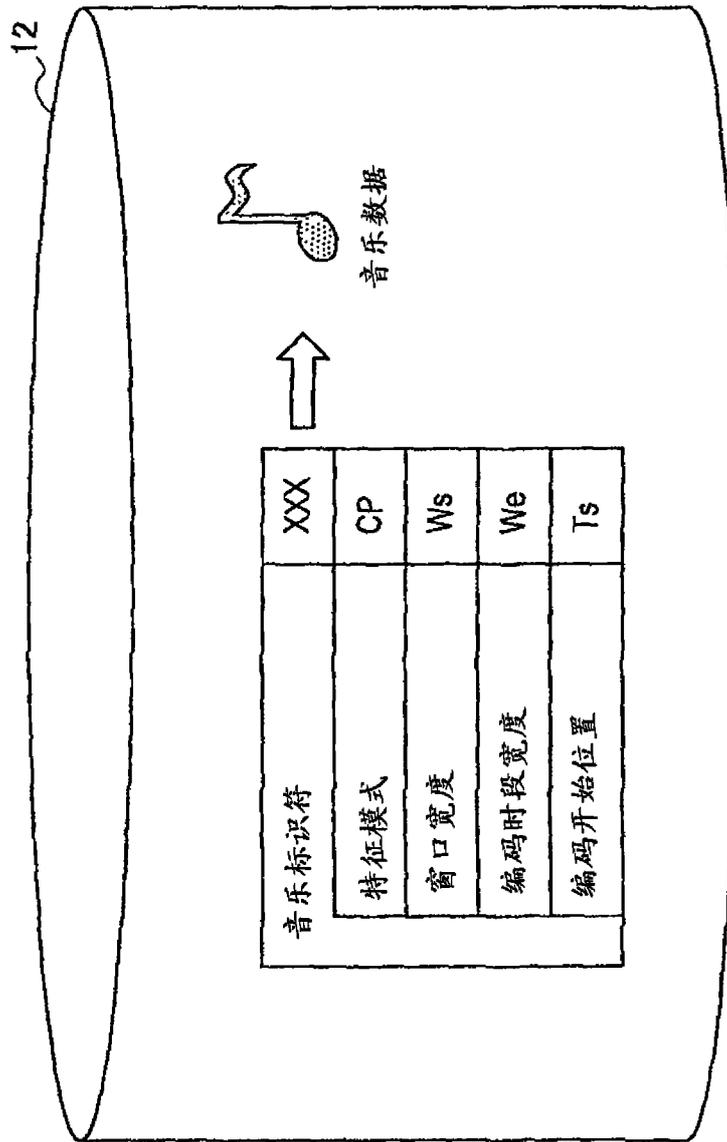


图 7

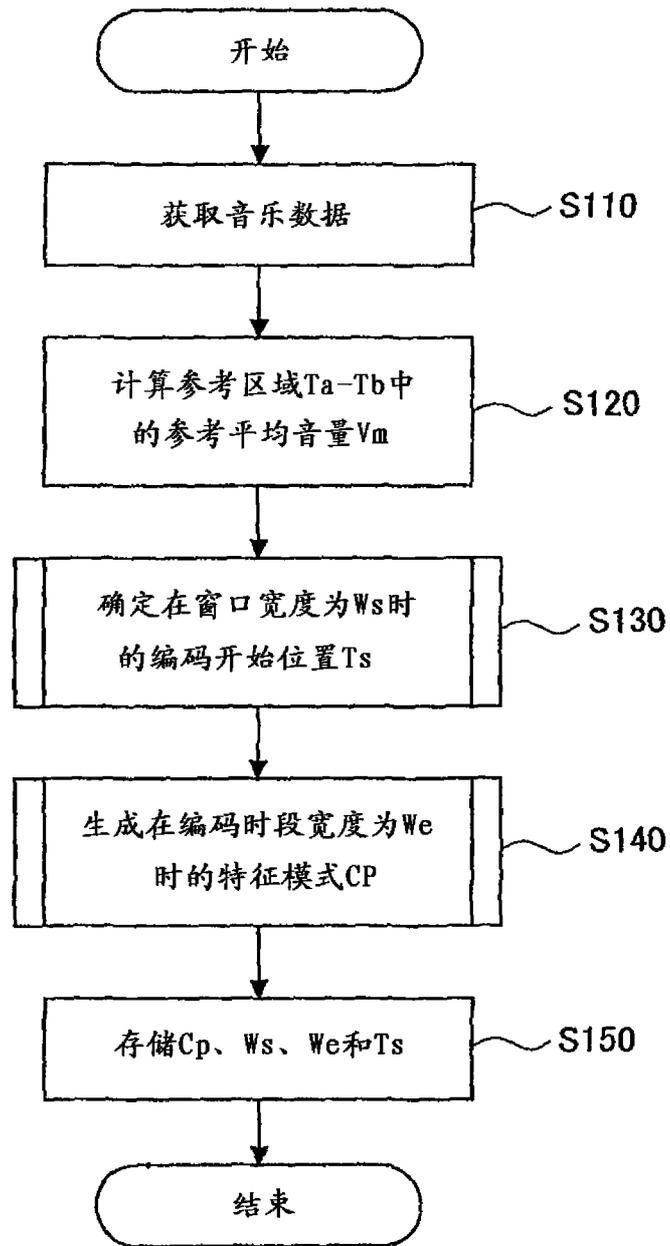


图 8

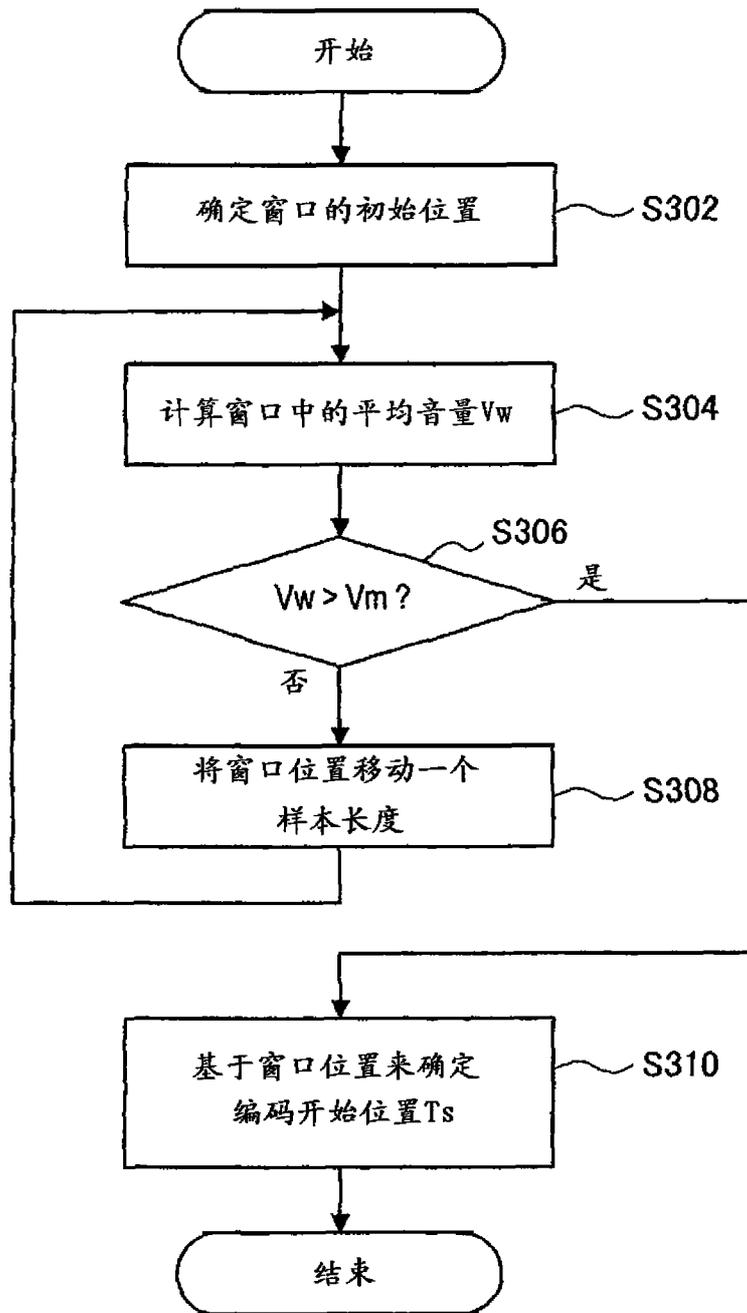


图 9

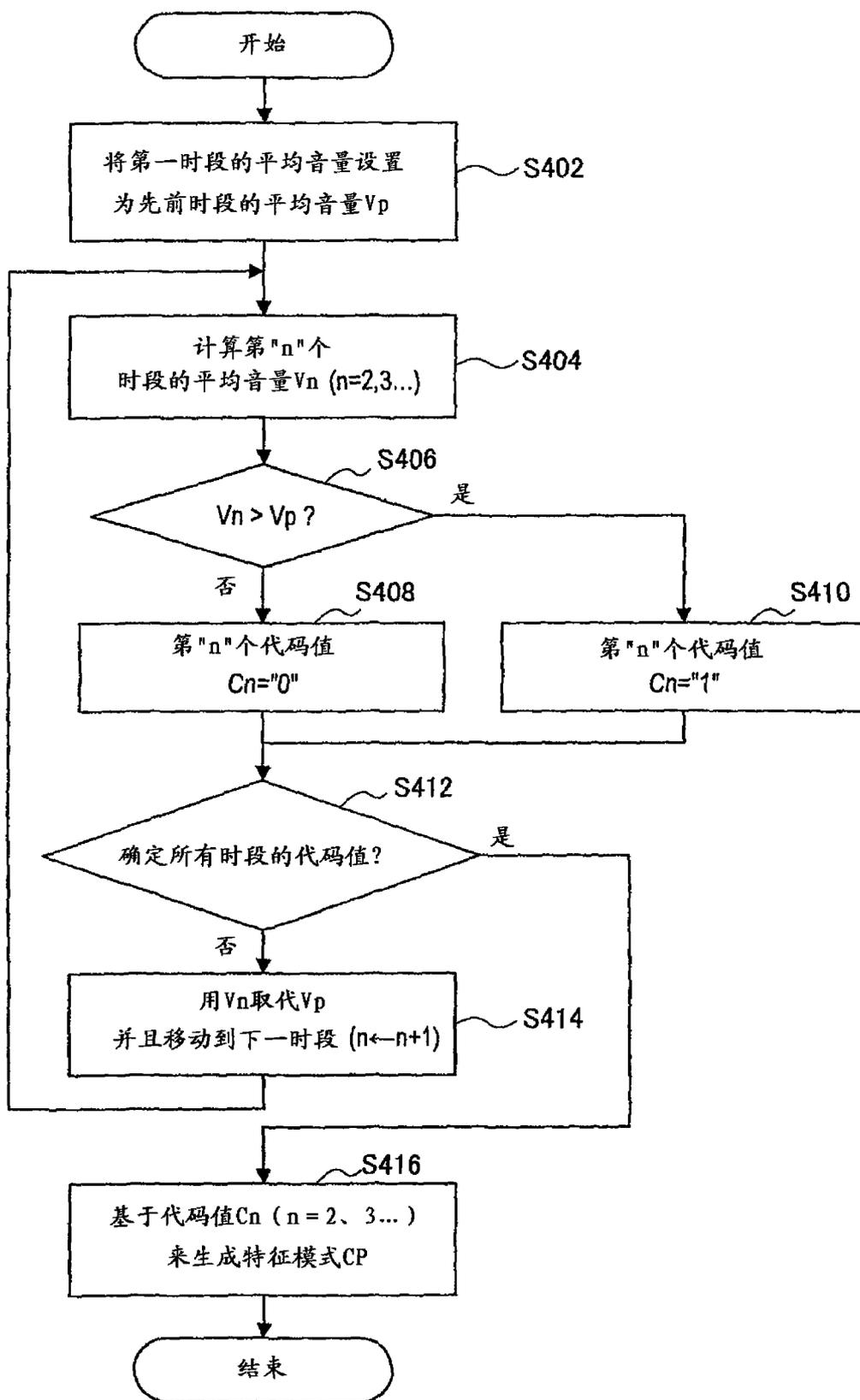


图 10

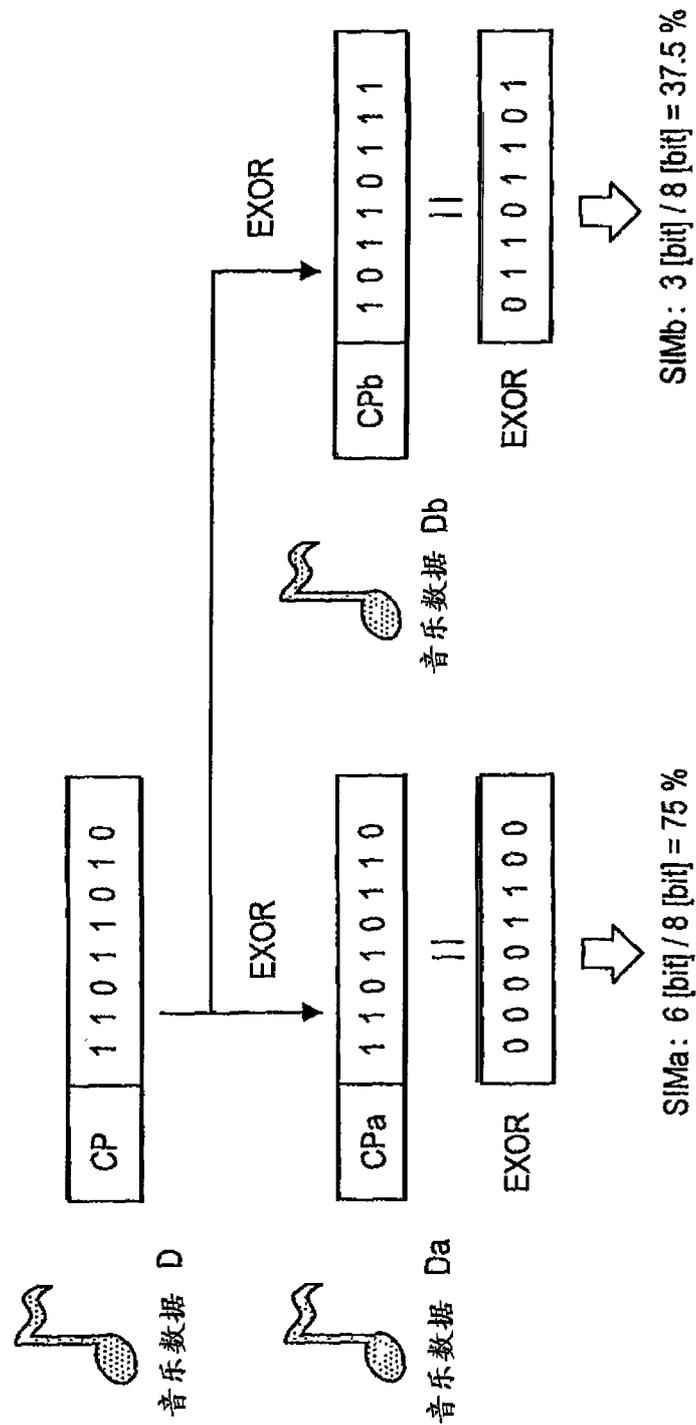


图 11

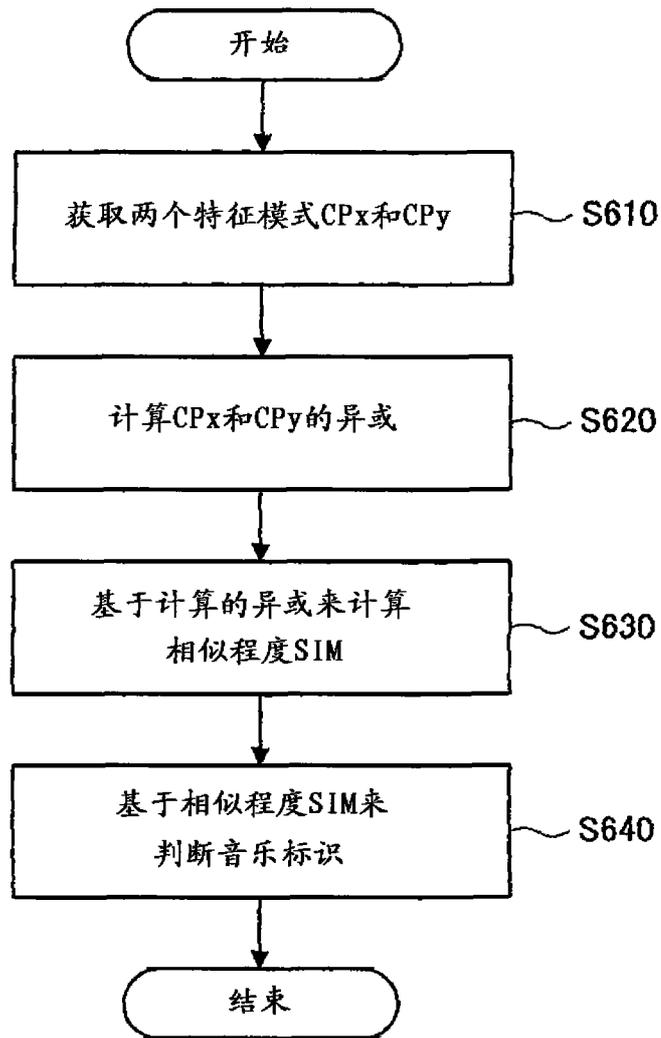


图 12

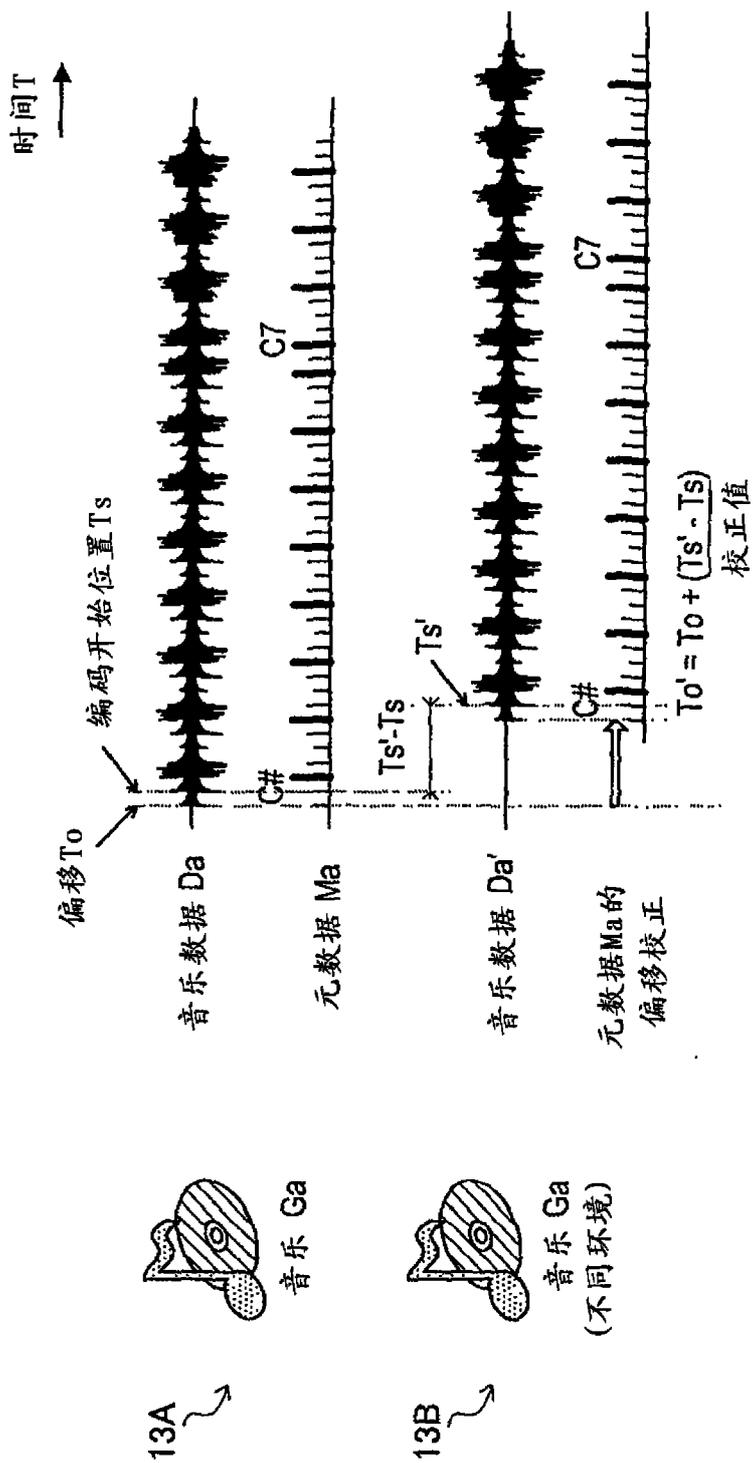


图 13

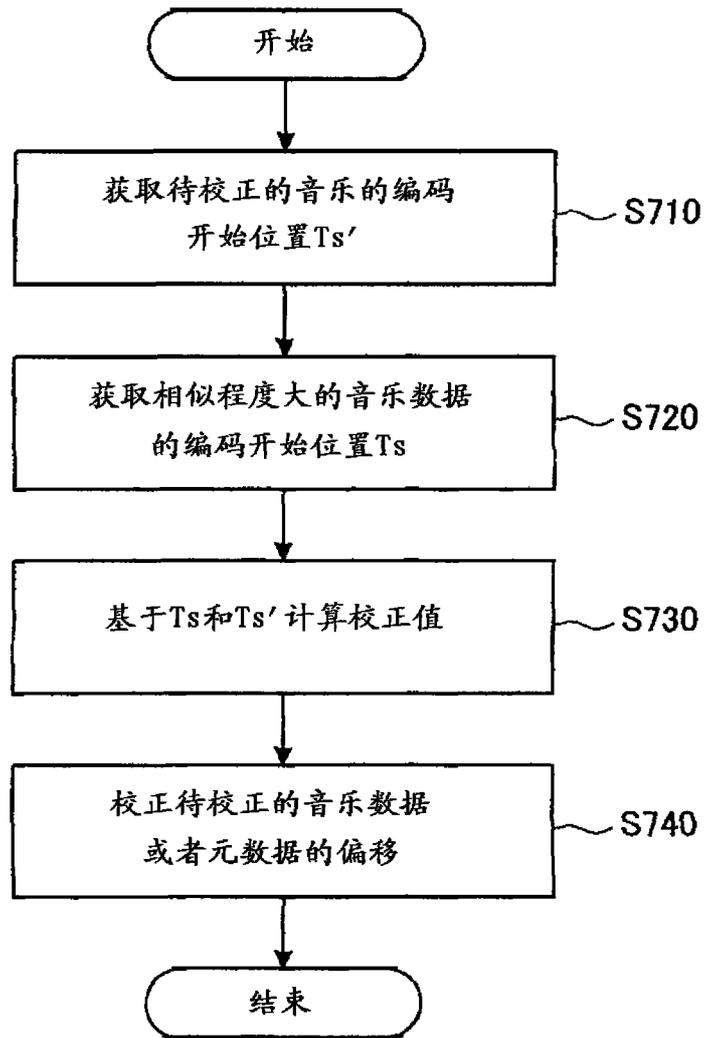


图 14

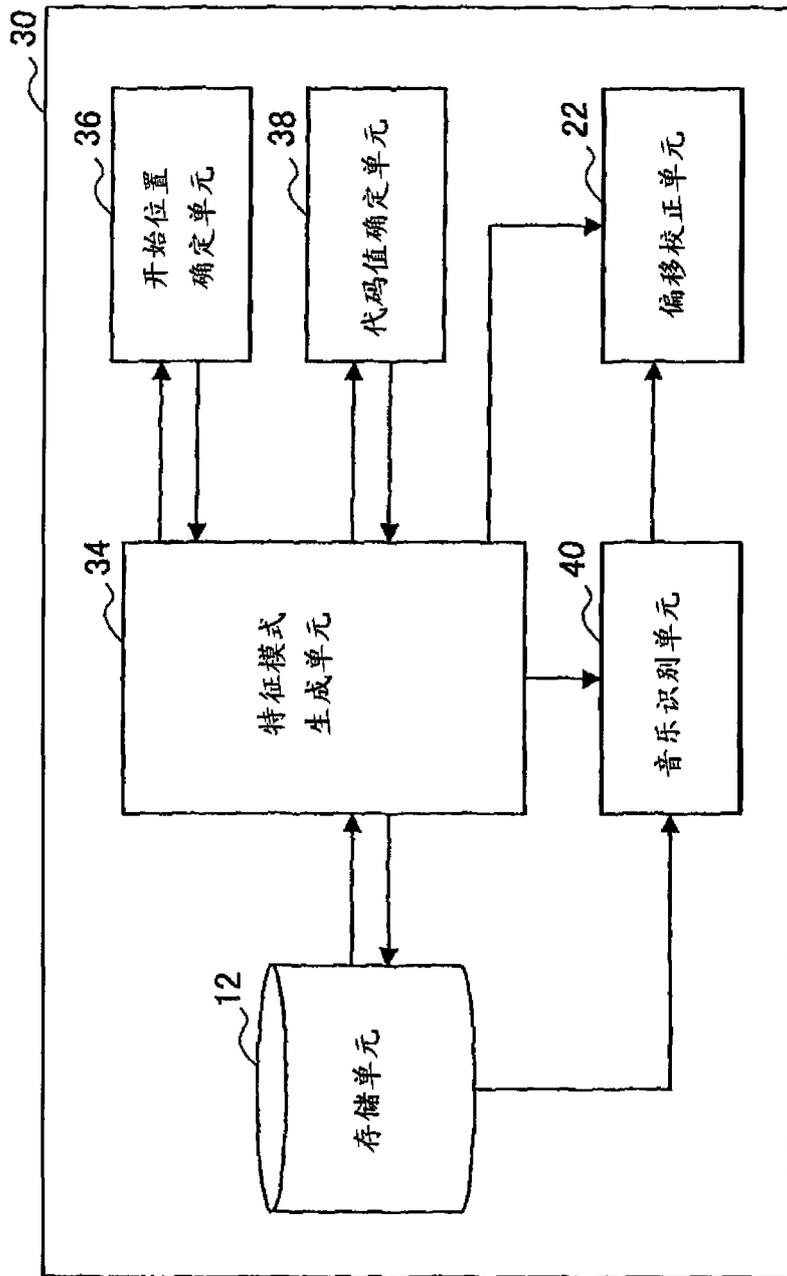


图 15

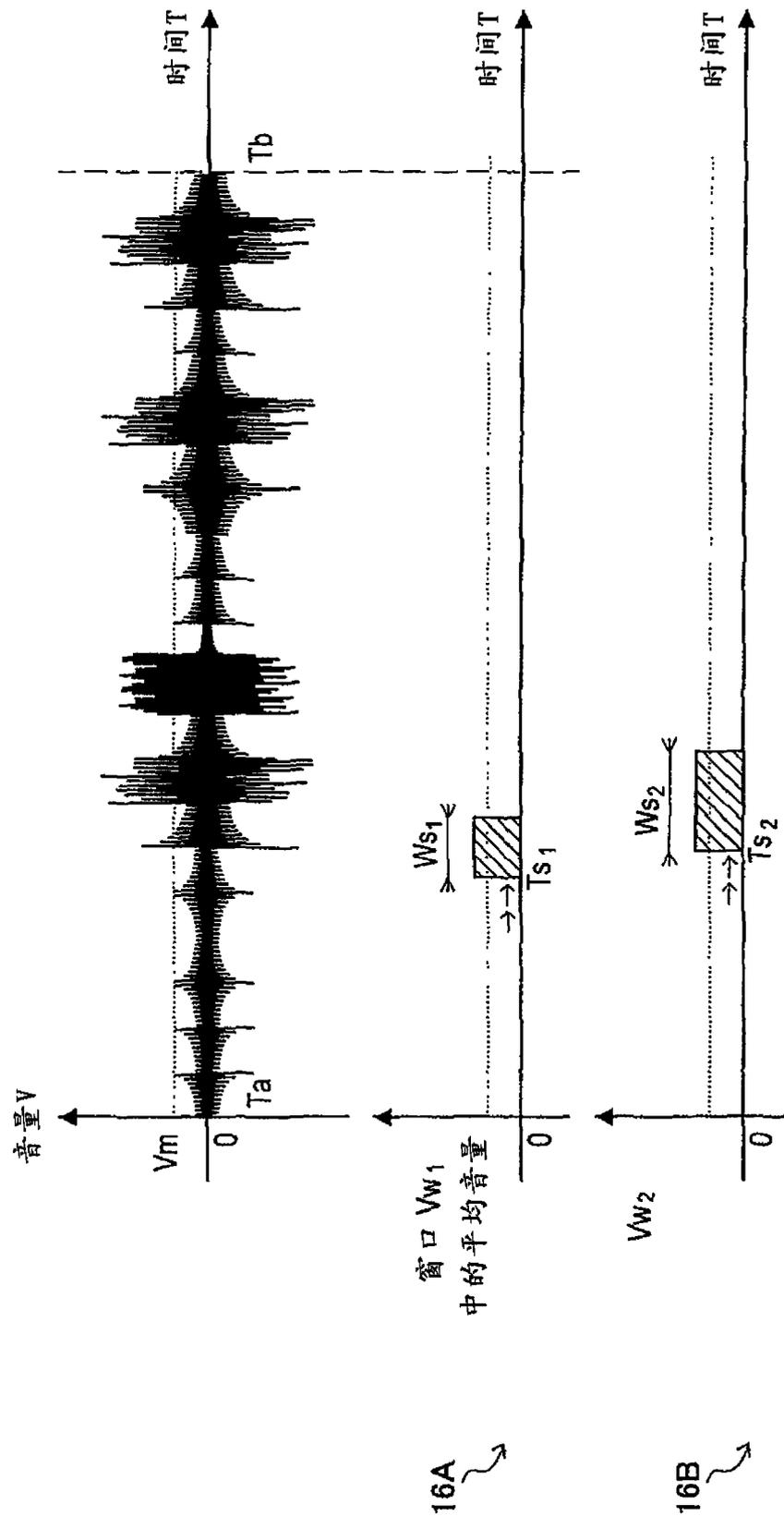


图 16

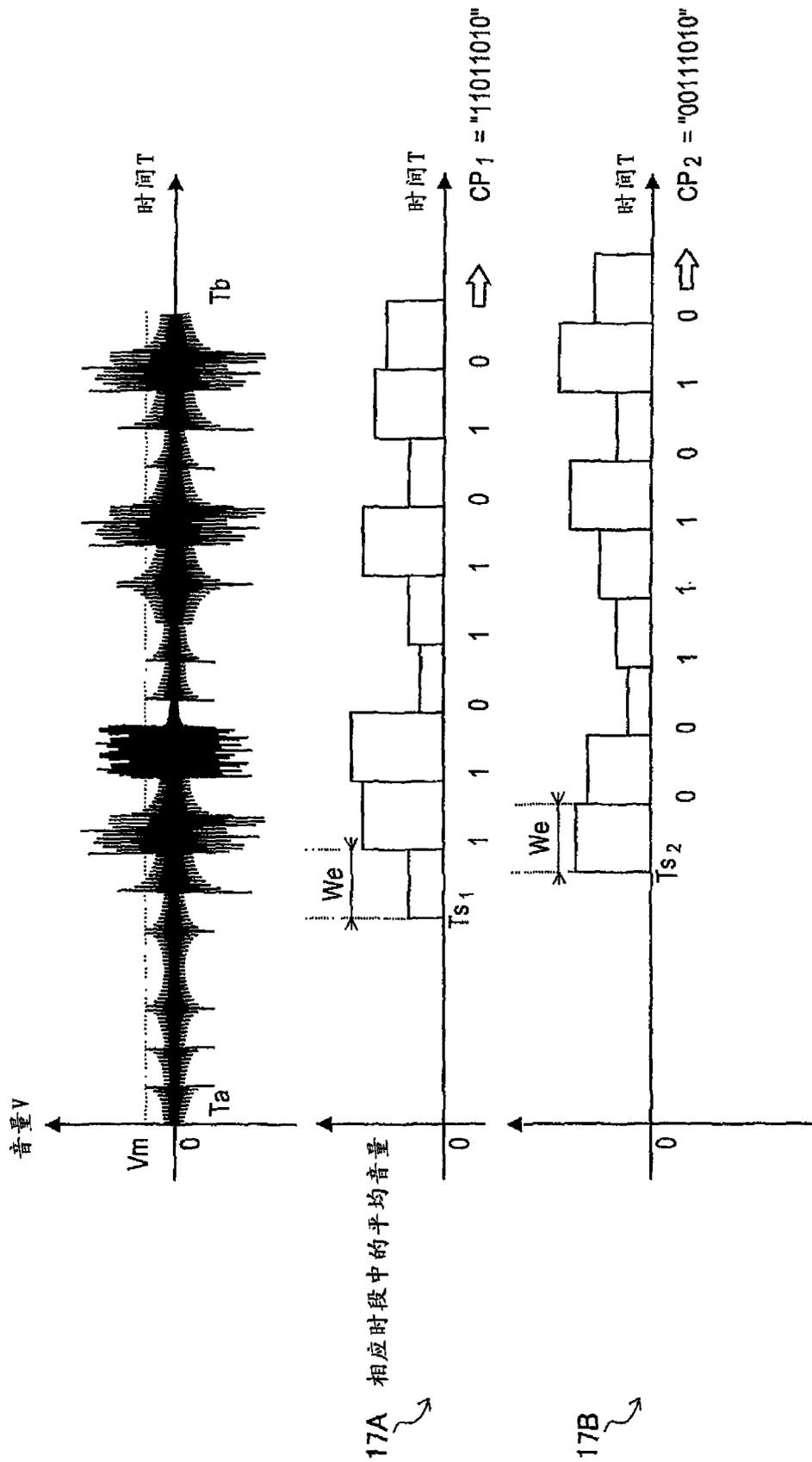


图 17

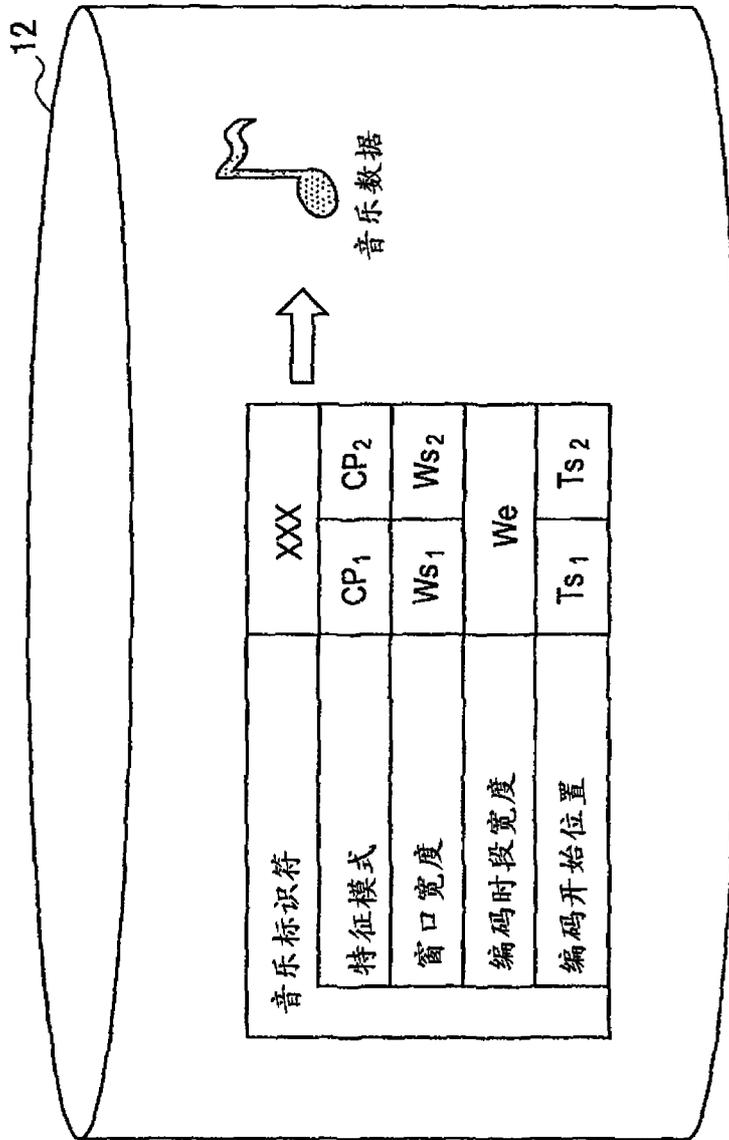


图 18

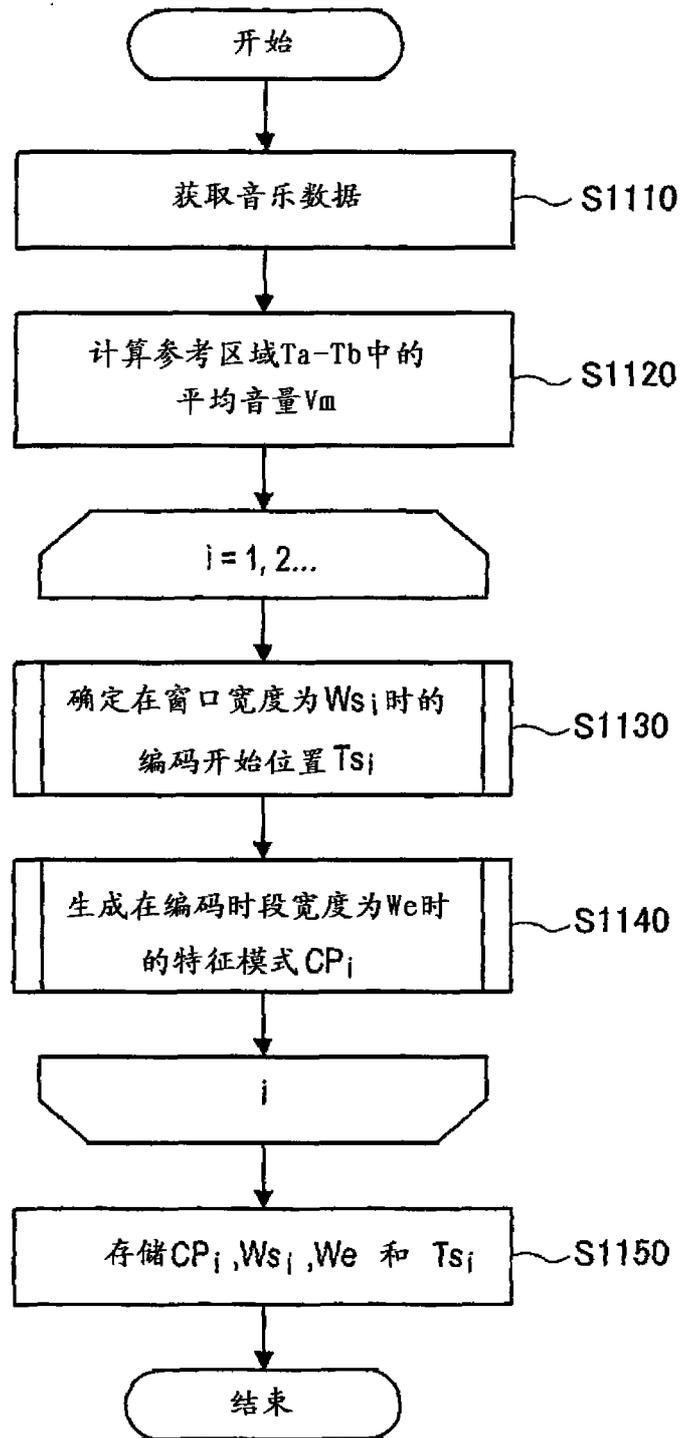


图 19

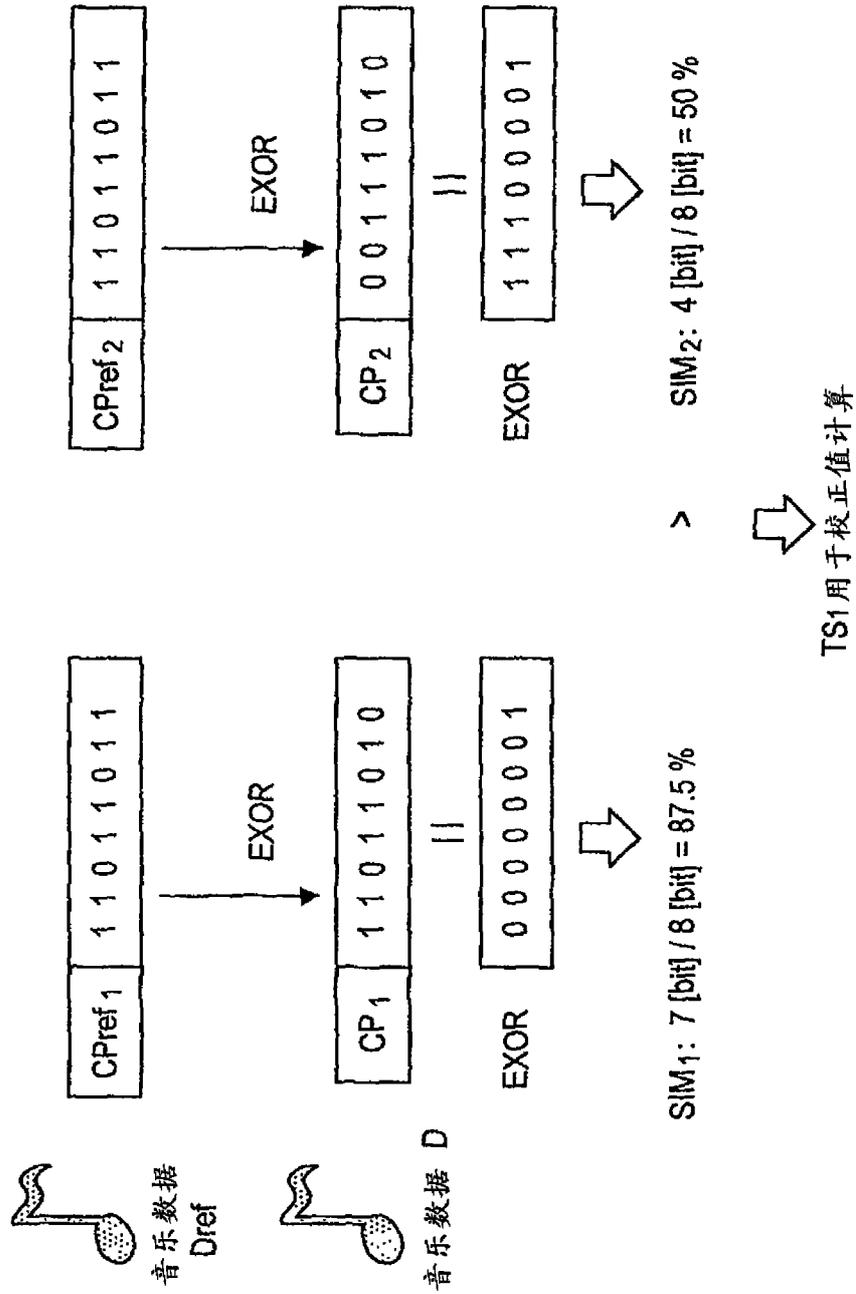


图 20

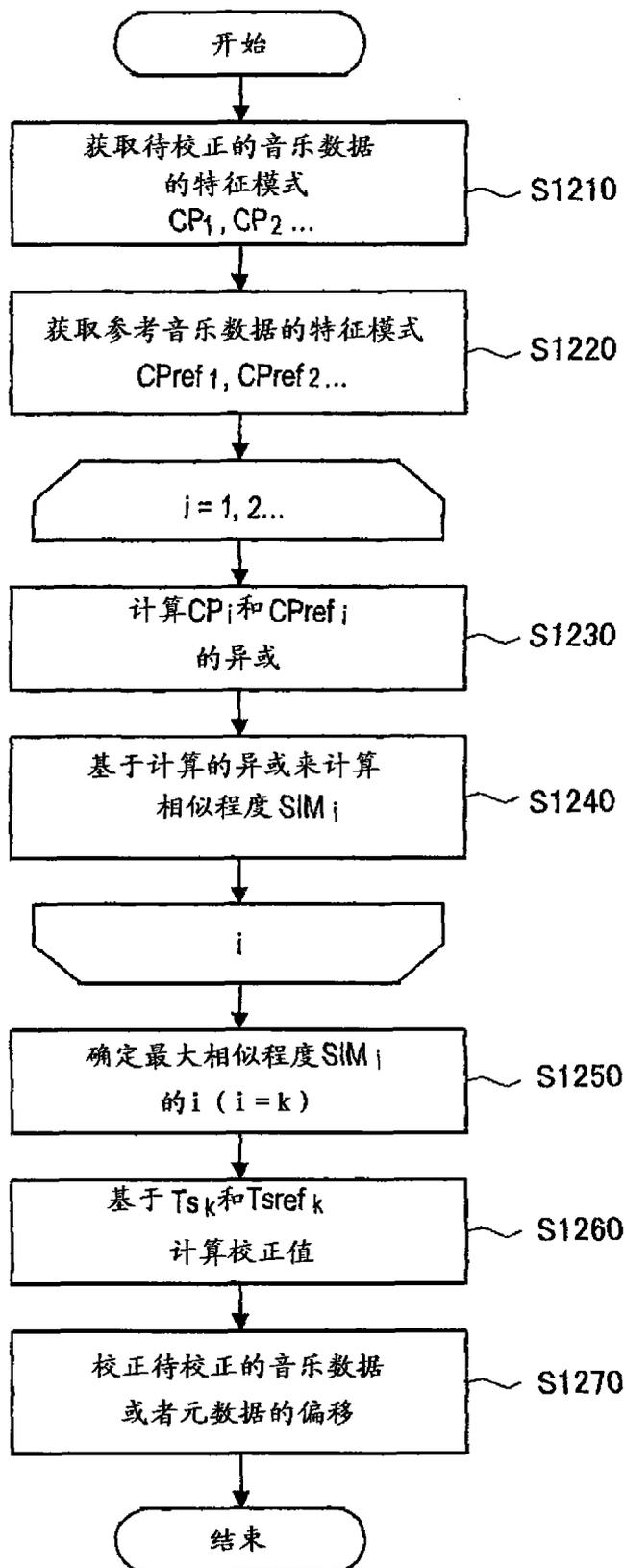


图 21

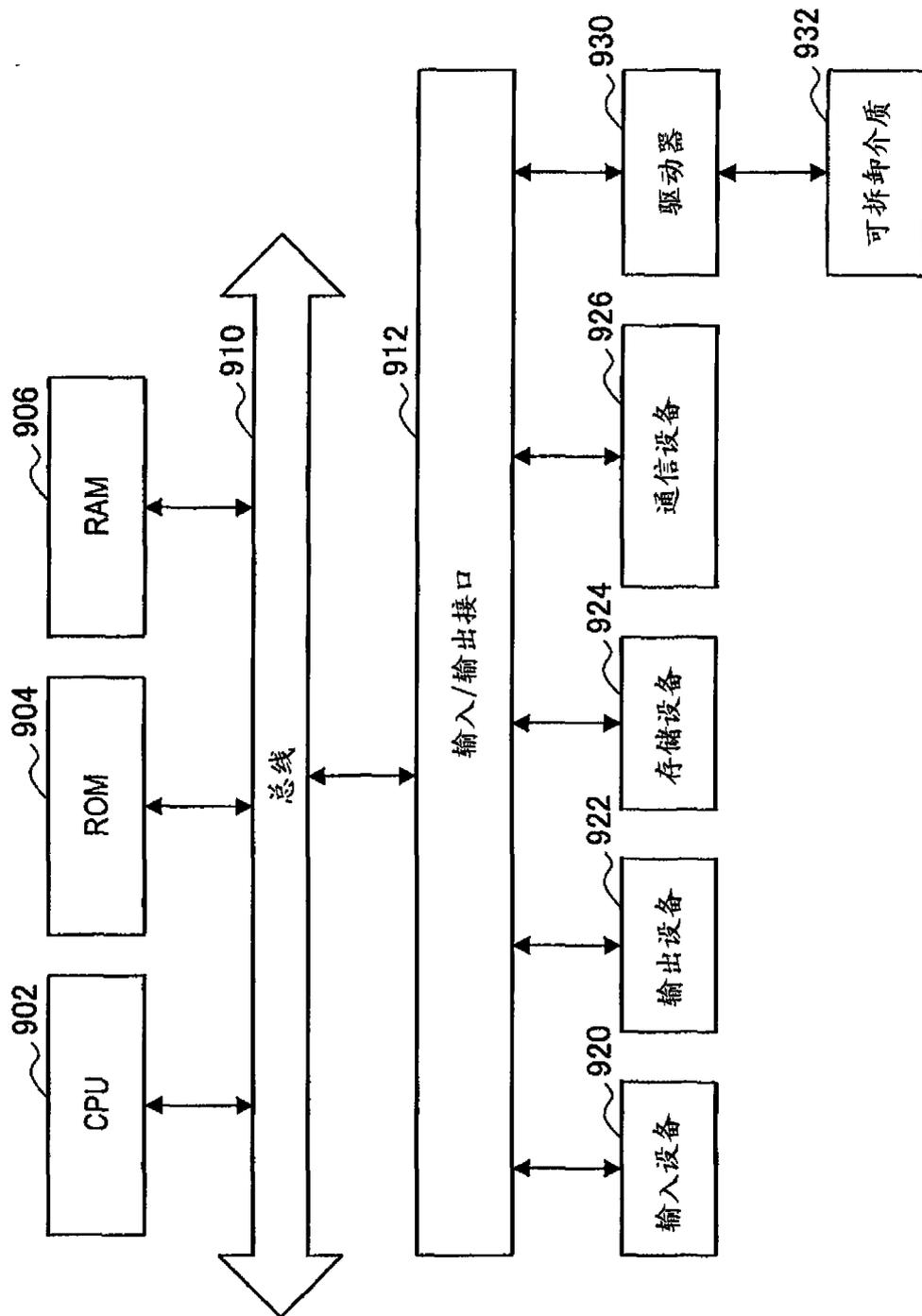


图 22