



(12) 发明专利申请

(10) 申请公布号 CN 101772949 A

(43) 申请公布日 2010.07.07

(21) 申请号 200880102117.3

(51) Int. Cl.

(22) 申请日 2008.07.17

H04N 1/21 (2006.01)

(30) 优先权数据

H04N 101/00 (2006.01)

11/834745 2007.08.07 US

(85) PCT申请进入国家阶段日

2010.02.05

(86) PCT申请的申请数据

PCT/US2008/008751 2008.07.17

(87) PCT申请的公布数据

W02009/020515 EN 2009.02.12

(71) 申请人 伊斯曼柯达公司

地址 美国纽约州

(72) 发明人 K·A·雅各比 C·W·洪辛格

T·J·默里 J·V·内尔逊

(74) 专利代理机构 中国专利代理(香港)有限公司
72001

代理人 王小衡 王忠忠

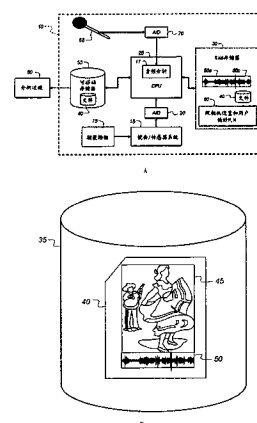
权利要求书 1 页 说明书 6 页 附图 7 页

(54) 发明名称

记录所捕获图像的音频元数据

(57) 摘要

一种在图像捕获期间记录音频元数据的方法包括:提供用于捕获静止或视频数字化的场景图像和记录音频信号的图像捕获设备;当该设备处于开机模式时将所述音频信号连续地记录在缓冲器中;以及由所述图像捕获设备启动静止图像或视频图像的捕获,并且将在静止图像或视频图像的捕获终止之前、期间和之后的时间内产生的音频信号存储为元数据。



1. 一种在图像捕获期间记录音频元数据的方法,包括:
 - a) 提供用于捕获静止或视频数字化的场景图像和记录音频信号的图像捕获设备;
 - b) 当该设备处于开机模式时将所述音频信号连续地记录在缓冲器中;以及
 - c) 由所述图像捕获设备启动静止图像或视频图像的捕获,并且将在静止图像或视频图像的捕获终止之前、或期间和之后的时间内产生的音频信号存储为元数据。
2. 权利要求 1 的方法,还包括在图像捕获设备中提供至少一个麦克风以及将该麦克风捕获的音频信号数字化以使得所记录的元数据音频信号被数字化。
3. 权利要求 1 的方法,其中音频信息被暂时存储在活动窗口存储缓冲器中。
4. 权利要求 1 的方法,还包括包含在视频图像捕获期间捕获的音频信号以及存储在存储器中的音频信号和在视频图像的捕获终止之后的预定时间期间产生的音频信号。
5. 权利要求 1 的方法,还包括为音频缓冲器提供默认持续时间。
6. 权利要求 1 的方法,还包括根据用户偏好调节要设置的音频缓冲器的持续时间。
7. 权利要求 6 的方法,还提供一种基于对音频信号的分析来确定捕获前音频缓冲器的持续时间和捕获后音频缓冲器的持续时间的自动模式。
8. 权利要求 1 的方法,其中音频信号被整体地存储在存储器中,并且存储器地址对待与图像数据相关联的音频元数据的开始和结尾进行标记。
9. 权利要求 7 的方法,还包括包含调节待与图像数据相关联的音频元数据的开始和结尾的存储器地址。
10. 权利要求 2 的方法,还包括提供与捕获的图像相关联的具有数字化图像和数字化音频元数据的图像文件。
11. 权利要求 4 的方法,还包括提供用于存储图像文件的可移动存储卡。
12. 权利要求 4 的方法,还包括分析音频元数据以提供对捕获的静止或视频图像的语义理解。
13. 权利要求 6 的方法,还包括提供音频元数据的书面文本。
14. 权利要求 6 的方法,还包括提供对出现在音频元数据中的环境声音的描述。
15. 权利要求 6 的方法,还包括提供音频元数据中的说话者的身份。
16. 权利要求 6 的方法,其中对音频元数据的分析发生在捕获设备内。
17. 权利要求 6 的方法,其中对音频元数据的分析发生在计算设备而不是捕获设备上。
18. 权利要求 6 的方法,还包括用从所述分析中获取的附加元数据来更新现有图像文件的元数据。
19. 权利要求 1 的方法,还包括在图像捕获之前存储音频信息。
20. 权利要求 1 的方法,还包括组合所存储的音频以形成音频剪辑。
21. 权利要求 1 的方法,其中在静止图像或视频图像的捕获终止之前、期间和之后的时间是可调节的。
22. 权利要求 20 的方法,还包括使用音频剪辑来提供对音频信息的语义理解,从而用于媒体搜索/检索。
23. 权利要求 1 的方法,还包括给突发捕获模式提供多个音频缓冲器以用于突发捕获序列中的每个静止图像。

记录所捕获图像的音频元数据

技术领域

[0001] 本发明通常涉及音频处理领域,并且更具体地涉及在相关联的静止或视频数字化图像的图像文件中嵌入音频元数据。

背景技术

[0002] 数字照相机往往包括视频捕获能力。另外,一些数字照相机具有用音频注解图像捕获数据的能力。音频波形往往被存储为数字编码的音频样本并且放置在文件格式的适当容器内,例如数字静止图像文件的元数据标签或者简单地作为视频文件或流中的(一个或多个)编码音频层。

[0003] 在消费电子学工业中已经有许多创新:将图像内容与声音进行结合。例如,伊斯曼·柯达公司在 US6496656B1 中教导了如何在硬拷贝印刷品中嵌入音频波形。另一个柯达专利 US6993196B2 教导了如何将音频数据作为非标准元数据存储于图像文件的结尾处。

[0004] Virage 公司具有一个专利 US6833865,该专利教导了关于一种用于实时嵌入的元数据提取的系统,所述元数据提取能够是与场景或音频相关的,只要音频已经存在于视听数据流中即可。该处理能够与捕获并行地或者顺序地进行。

[0005] US7113219B2 是惠普专利,其教导了使用按钮上的第一位置来捕获音频以及使用第二位置来捕获图像。

[0006] 尽管这样的音频信息驻留在图像或视频文件中以用于回放目的,但是该音频除了允许在以后查看文件时回放声音之外没有其它目的。当前,没有机制用于在捕获时或在以后自动地捕获与数字图像或视频捕获同时的音频事件以便后续分析进行理解、组织、分类或搜索/检索。

发明内容

[0007] 简而言之,依据本发明,提供了一种在图像捕获期间记录音频元数据的方法,包括:

[0008] a) 提供用于捕获静止或视频数字化的场景图像和记录音频信号的图像捕获设备;

[0009] b) 当该设备处于开机模式时连续地记录所述音频信号;以及

[0010] c) 由所述图像捕获设备启动静止图像或视频图像的捕获,并且将在静止图像或视频图像的捕获终止之前、期间和之后的时间内产生的音频信号存储为元数据。

[0011] 本发明自动地将音频元数据与图像捕获相关联。而且,本发明自动地将同时发生的音频信息的预定段与图像或图像的视频序列相关联。

[0012] 要理解,如本发明的该说明书中使用的短语“图像捕获”、“捕获的图像”、“图像数据”涉及静止图像捕获以及运动图像捕获,如在视频中。术语“静止图像捕获”和“视频捕获”或者其变型在需要时将用来描述不同的静止或运动捕获情况。

[0013] 本发明的优点源于如下事实:在图像捕获之前、期间和之后被捕获的记录的音频

信息提供场景的情境以及有用的元数据,其能够被分析以语义理解所捕获的图像。依据本发明,一个过程将音频信息的不断更新的活动窗口与所捕获的图像相关联,以允许用户有不必要经过驱动按钮或开关来主动地启动音频捕获的自由。用户所需的物理动作是启动图像或视频捕获事件。音频信号与(多个)图像的关联以及音频信息的活动窗口的管理由设备的电子器件自动地处理并且对用户是完全透明的。

[0014] 通过查看以下优选实施例的详细描述和所附权利要求并且参照附图,将更清楚地理解和明白本发明的这些以及其它方面、目标、特征和优点。

[0015] 本发明包括这些优点:存储于存储器中的开机模式中的音频的连续捕获允许捕获能够用于对图像数据的语义理解的更多信息,以及在查看图像数据时经过音频的回放来增强用户体验。在图像捕获时,来自静止和视频捕获之前的一段时间、静止和视频捕获期间以及静止和视频捕获之后的一段时间的音频样本作为元数据被自动地存储在图像文件中以用于以后的语义分析。

附图说明

[0016] 图 1a 是描述发明的实施例的框图;

[0017] 图 1b 示出含有图像和音频数据的多媒体文件;

[0018] 图 2a 是描述代表性摄影环境的草图,含有照相机用户、对象、场景以及其它产生环境中声音的目标;

[0019] 图 2b 是使用发明的优选实施例说明在典型使用情况下发生的高级事件的流程图;

[0020] 图 3a 是示出作为与静止图像捕获情况交叠的时变信号的数字化音频信号波形的详细图;

[0021] 图 3b 是特定于视频捕获情况的数字化音频信号波形的详细图;以及

[0022] 图 4 是用于分析所记录的音频信号的图 1a 所示的分析过程的框图。

具体实施方式

[0023] 在以下描述中,本发明将在其优选实施例中描述为数字照相机设备。本领域技术人员将容易意识到等效发明还能够存在于其它实施例中。

[0024] 图 1a 示出了数字照相机设备 10 的示意图。数字照相机设备 10 含有用于图像捕获的照相机镜头及传感器系统 15。图像数据 45(参见图 1b)能够为单独的静止图像或者如视频中的一系列图像。这些图像数据由专用的图像模拟-数字转换器 20 量化并且计算机 CPU 25 处理该图像数据 45 并将其编码为数字多媒体文件 40 以存储在内部存储器 30 或可移动存储器模块 35 中。内部存储器 30 还为捕获前缓冲的音频信号 55a 和捕获后缓冲的音频信号 55c 以及为照相机设置和用户偏好 60 提供足够的存储空间。另外,数字照相机设备 10 含有麦克风 65,麦克风 65 记录场景的声音或者记录语音以用于其它目的。麦克风 65 所生成的电信号由专用的音频模拟-数字转换器 70 数字化。数字音频信号 175 被存储在内部存储器 30 中作为捕获前缓冲的音频信号 55a 和捕获后缓冲的音频信号 55c。

[0025] 图 1b 示出了含有数字多媒体文件 40 的可移动存储器模块 35(例如 SD 存储卡或存储棒)的图。文件含有先前提到的图像数据 45 以及相伴的音频剪辑 50。

[0026] 在图 2a 所示的优选实施例的共用情况内能够更好地理解图 1a 中所描述的各种部件的操作,其中图 2a 描述了代表性摄影环境。参照图 2a,带有数字照相机设备 10 的摄影师 90 与环境 85 中的对象 100 进行言语交互。环境 85 被定义为其中目标对数字照相机设备 10 是可见的或可听到的空间。摄影师 90 和对象 100 的各自言语 95 和 105 能够是对话的一部分,或者能够是由对象 100 或摄影师 90 如以讲述或注解的方式单向地产生。摄影场景 130 被定义为数字照相机设备 10 的光学视场。在环境 85 中能够存在由其它场景相关目标 110 所产生的其它场景相关环境声音 115。在图 2a 的情况下,场景相关目标 110 是在摄影场景 130 内的音乐家。来自被示为飞机的非场景相关目标 120 的非场景相关环境声音 125 对麦克风 65 是可听到的并且因此是数字照相机设备 10 感测的环境 85 的一部分,然而它们不是摄影场景 130 的一部分。图 2a 还示出了总声音 135,其被定义为入射到麦克风 65 上的环境 85 内的所有声源的总和。

[0027] 图 2b 是涉及捕获图 2a 所示的摄影场景 130 的静止图像的事件序列的流程图。参照图 2b,数字照相机设备 10 开机或唤醒步骤 140 示出了通过打开电源来激活数字照相机设备 10 或者以其它方式从休眠或待机模式中唤醒。这个步骤很重要,因为在音频信号缓冲步骤 145 中数字照相机设备 10 立即开始将麦克风 65 产生的数字音频信号 175(参见图 3a)存储为捕获前缓冲的音频信号 55a。音频信号缓冲步骤 145 允许摄影师 90 在图像捕获事件 150 之前进行与摄影场景 130 或环境 85 中的对象 100 或其它属性的进行对话或者描述所述对象 100 或其它属性。同时,还可能存在麦克风 65 感测的其它非言语声音,诸如先前讨论的场景相关环境声音 115 或其它非场景相关环境声音 125,其能够给随后的图像捕获事件 150 添加额外的情境。重要的是注意,在音频信号缓冲步骤 145 中麦克风 165 和音频模拟-数字转换器 70 记录环境 85 中存在的总声音 135。在图像捕获事件 150 中,摄影师 90 按下捕获按钮 75(参见图 1a),这就启动捕获摄影场景 130 的图像数据 45。在继续的音频信号缓冲步骤 155 中数字照相机设备 10 继续记录来自环境 85 的总声音 135 达照相机设置和用户偏好 60 中指定的额外时间段。

[0028] 在这一点上,图 2b 的流程图更详细地示出了在音频信号缓冲步骤 145 到继续的音频信号缓冲步骤 155 期间所发生的情况。参照图 3a,示出了由麦克风 65 拾取的作为数字音频信号 175 表示的总声音 135、以及相关的时间线 180。如先前所陈述的,在音频信号缓冲步骤 145 中,总声音 135 被连续地存储为捕获前缓冲的音频信号 55a。捕获前缓冲的音频信号 55a 存储 N 秒的音频信息,如时间线 180 上由时间线 180 上的“ $t = -N$ ”时间标记 185 所示。“ $t = -N$ ”时间标记 185 指明捕获前缓冲的音频信号 55a 的时间上的起点。这个捕获前缓冲的音频信号 55a 以“活动窗口”的方式进行连续更新,其中在时间线 180 上最老的样本在“ $t = -N$ ”时间标记 185 处溢出缓冲器的末端而当前的音频样本填充在“ $t_0 = 0$ ”时间标记 190a 处的缓冲器的前端。“ $t_0 = 0$ ”时间标记 190a 代表数字照相机设备 10 打开并正在收听环境 85 中存在的总声音 135 时的实时当前时刻。捕获前缓冲的音频信号 55a 能够被认为是在从“ $t = -N$ ”时间标记 185 跨越到“ $t_0 = 0$ ”时间标记 190a 的 FIFO(先进先出)样本向量中不断更新的声音的活动窗口。

[0029] 回头参照图 2b,图像捕获事件 150(即,摄影师 90 按下捕获按钮 75)与捕获前缓冲的音频信号 55a 的填入(population)的完成相一致。在“ $t_0 = 0$ ”时间标记 190a 处发生的图像捕获事件 150 的时候,继续的音频信号缓冲步骤 155 示出了数字音频信号 175 继续

填充捕获后音频数据缓冲器 55c 达额外的 M 秒,如时间线 180 上的“ $t = +M$ ”时间标记 195 所示。在静止图像捕获的情况下,理想化的是图像捕获事件 150(参见图 3a)捕获时间上的无穷小时刻,然而图像捕获事件实际上跨越了快门的持续时间或传感器的积分时间。例如,数字照相机设备 10 的曝光时间可以在照相机设置和用户偏好 60 中被设置为 1/20 秒。在一秒的这个分数期间的音频以无缝的方式被保存以使数字音频信号 175 从“ $t = -N$ ”时间标记 185 跨越到“ $t = +M$ ”时间标记 195。在音频剪辑形成步骤 157 中捕获前缓冲的音频信号 55a 和捕获后缓冲的音频信号 55c 被组合以形成音频剪辑 50(参见图 3a)。

[0030] 图 3b 示出了特定于视频捕获情况的音频波形的图,其中总声音 135(参见图 2a)被记录同时数字照相机设备 10 的照相机镜头及传感器系统 15(参见图 1a)将图像数据 45(参见图 1b)记录为视频帧。图像数据 45 被捕获同时数字音频信号 175 继续被记录和存储为视频流 55b' 的音频部分达图像捕获事件 150 的持续时间;例如达额外的 T 秒,如由从“ $t_0 = 0$ ”时间标记 190a 到完成图像捕获事件 150 后的“ $t_1 = +T$ ”时间标记 190b 的时间跨度所示。视频捕获前缓冲的音频信号 55a'、视频流 55b' 的音频部分以及视频捕获后缓冲的音频信号 55c' 被合并以形成与图像捕获事件 150 相关联的音频剪辑 50。

[0031] 回头参照图 2b,在视频捕获的情况下,音频剪辑形成步骤 157 组合视频捕获前缓冲的音频信号 55a'、视频流 55b' 的音频部分以及捕获后缓冲的音频信号 55c'(参见图 3b)。音频剪辑存储步骤 160 将音频剪辑 50 存储为数字多媒体文件 40 的一部分。在语义分析步骤 165 中,音频剪辑 50 通过语义分析过程 80(参见图 1a)经历进一步的分析。最后,增强的用户体验步骤 170 示出音频剪辑 50 能够被用于增强的用户体验。例如,音频剪辑 50 能够在查看图像数据时进行简单的回放。另外,作为语义分析步骤 165 的结果的从音频剪辑 50 中搜集的信息构成新的元数据 205(参见图 4)并且能够例如被用来增强基于语义的媒体搜索和检索。

[0032] 图 4 是用于语义分析步骤 165(参见图 2b)的音频数据分析的更详细的框图。语义分析过程 80 在本发明的优选实施例中是语音到文本操作 200,语义分析过程 80 将出现在音频剪辑 50 中的言辞转换成新的元数据 205。能够进行其它分析,例如检查音频剪辑 50 以帮助对捕获位置和条件的语义理解,检测目标或人物的出现或身份。在优选实施例中,新的元数据 205 采取一系列识别的关键字的形式,或者其能够是一列短语或语音串。新的元数据 205 通过将元数据写到文件操作 210 而与数字多媒体文件 40 相关联。

[0033] 回头参照图 3a 和 3b,捕获前缓冲的音频信号 55a(视频捕获前缓冲的音频信号 55a')和捕获后缓冲的音频信号 55c(视频捕获后缓冲的音频信号 55c')的持续时间具有默认值并且在照相机设置和用户偏好 60 中(参见图 1a)是用户可调节的,所述照相机设置和用户偏好 60 被存储在内部存储器 30 中。例如,捕获前缓冲的音频信号 55a 默认持续时间能够在照相机设置和用户偏好 60 中被预置为 $N = 10$ 秒,而捕获后缓冲的音频信号 55c 默认持续时间能够在照相机设置和用户偏好 60 中被预置为 $M = 5$ 秒。缓冲器的持续时间是任意的并且在需要更多或更少时间的情况下是用户可调节的。

[0034] 如果在捕获后缓冲的音频信号 55c 仍在用音频样本填入自身的过程中启动另一捕获事件 150(这将是处于突发模式捕获的情况),则能够支持内部存储器 30 中的多个缓冲器(参见图 1a)。

[0035] 获得等效音频剪辑 50 的另一方法将是把全部数字音频信号 175(参见图 3a、3b)

存储在数字照相机设备 10 的内部存储器 30 中,假设内部存储器 30 的存储容量足够的话。在用户希望捕获图像数据 45(参见图 1b)的那个时候,用户按下捕获按钮 75(参见图 1a)以启动在“ $t_0 = 0$ ”时间标记 190a 处发生的捕获事件 150(参见图 3a、3b)。在捕获事件 150 的初始“ $t_0 = 0$ ”时间标记 190a 处,位于在“ $t_0 = 0$ ”时间标记之前 N 秒的“ $t = -N$ ”时间标记 185 处的时移指针定义音频剪辑 50 的开始,一旦捕获后缓冲的音频信号 55c 完成,所述音频剪辑 50 将包括从“ $t = -N$ ”时间标记 185 到“ $t = +M$ ”时间标记 195 的音频样本。

[0036] 除了具有预置的时间长度来捕获在图像捕获事件之前和之后的音频之外,还可能要慎重的是在‘切断数字音频信号’之前实时地分析数字音频信号 175 以确定音频的连续性。例如,发生在数字照相机设备 10 的计算机 CPU 25 内的连续音频分析过程 17(参见图 1a)能够实时地分析数字音频信号 175(参见图 3a、3b)并且确定适当的位置以开始和结束音频剪辑。例如,如果数字音频信号 175 包括口头独白,则通过自动调节“ $t = -N$ ”时间标记 185 将保存更长或更短的捕获前缓冲的音频信号 55a,或者通过自动调节“ $t = +M$ ”时间标记 195 将保存更长或更短的捕获后缓冲的音频信号 55c,以便维持数字音频信号 175 的连续性。基于音频连续性或响度阈值找出数字音频信号 175 中的适合 (convenient) 中断允许系统适当地剪辑数字音频信号 175,而‘固定’时间可能在字中间切断数字音频信号 175。换句话说,人们可能期望如果数字音频信号 175 下降到阈值之下预定时间量则终止数字音频信号 175 捕获,因而为声音不重要时的那些情况节省文件空间。相反,可能存在太多的噪声以致声音对语义或重复使用等而言是无用的。音频分析过程 17 将采用针对音频可用性的阈值并抛弃任何响亮的、非可辨别的或连续的噪声。

[0037] 部件列表

[0038] 10 数字照相机设备

[0039] 15 照相机镜头及传感器系统

[0040] 17 音频分析过程

[0041] 20 图像模拟 - 数字转换器

[0042] 25 计算机 CPU

[0043] 30 内部存储器

[0044] 35 可移动存储器模块

[0045] 40 数字多媒体文件

[0046] 45 图像数据

[0047] 50 音频剪辑

[0048] 55a 捕获前缓冲的音频信号

[0049] 55a' 视频捕获前缓冲的音频信号

[0050] 55b' 视频流的音频部分

[0051] 55c 捕获后缓冲的音频信号

[0052] 55c' 视频捕获后缓冲的音频信号

[0053] 60 照相机设置和用户偏好

[0054] 65 麦克风

[0055] 70 音频模拟 - 数字转换器

[0056] 75 捕获按钮

- [0057] 80 语义分析过程
- [0058] 85 环境
- [0059] 90 摄影师
- [0060] 95 摄影师的言语 / 声音
- [0061] 100 对象
- [0062] 105 对象的言语 / 声音
- [0063] 110 场景相关目标
- [0064] 115 场景相关环境声音
- [0065] 120 非场景相关目标
- [0066] 125 非场景相关环境声音
- [0067] 130 摄影场景
- [0068] 135 总声音
- [0069] 140 设备开机或唤醒步骤
- [0070] 145 音频信号缓冲步骤
- [0071] 150 图像捕获事件（静止或视频）
- [0072] 155 继续的音频信号缓冲步骤
- [0073] 157 音频剪辑形成步骤
- [0074] 160 音频剪辑存储步骤
- [0075] 165 语义分析步骤
- [0076] 170 增强的用户体验步骤
- [0077] 175 数字音频信号
- [0078] 180 时间线
- [0079] 185t = -N 时间标记
- [0080] 190a t0 = 0 时间标记
- [0081] 190b t1 = T 时间标记
- [0082] 195t = +M 时间标记
- [0083] 200 语音到文本操作
- [0084] 205 新的元数据
- [0085] 210 将元数据写到文件操作

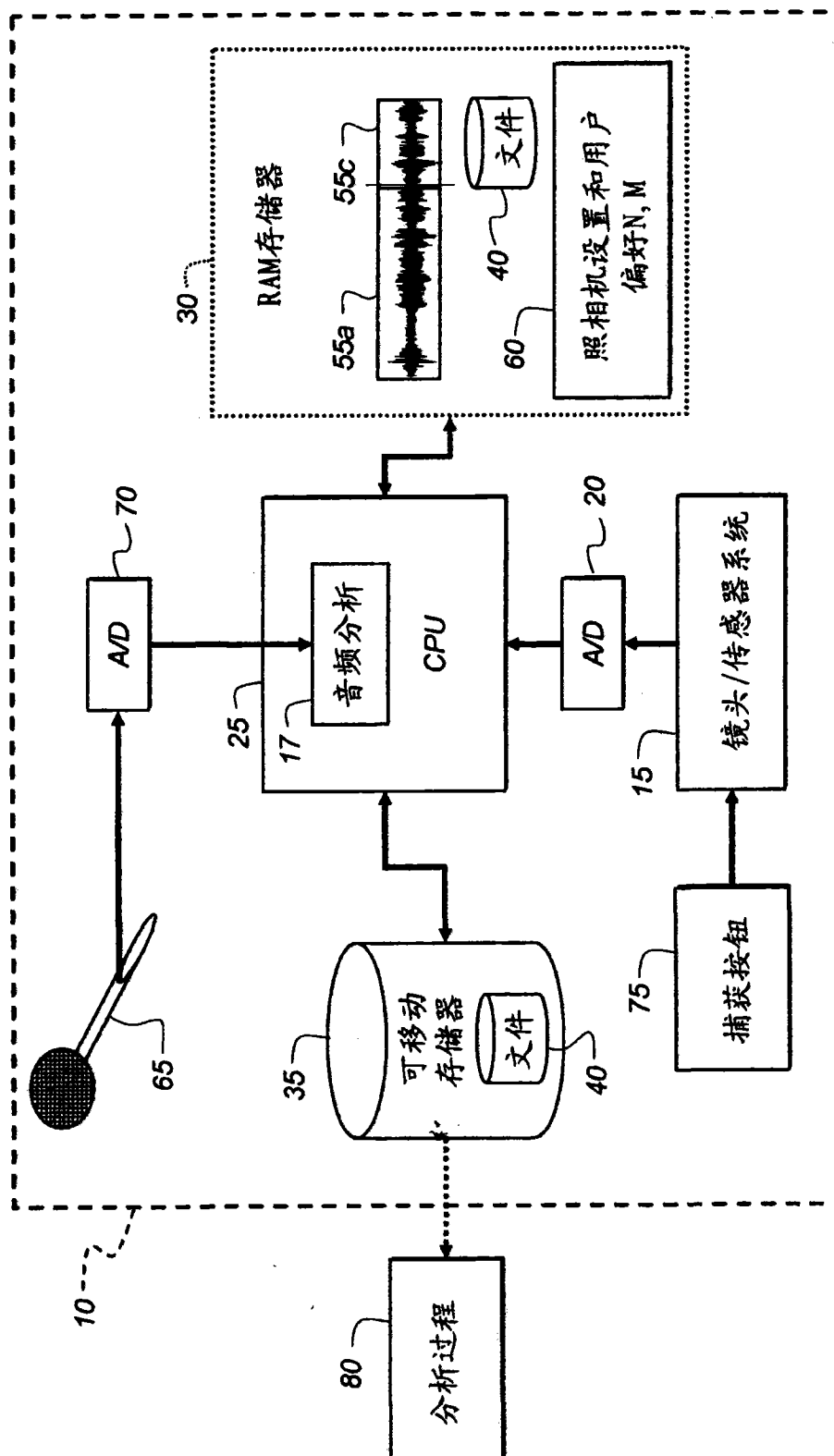


图 1A

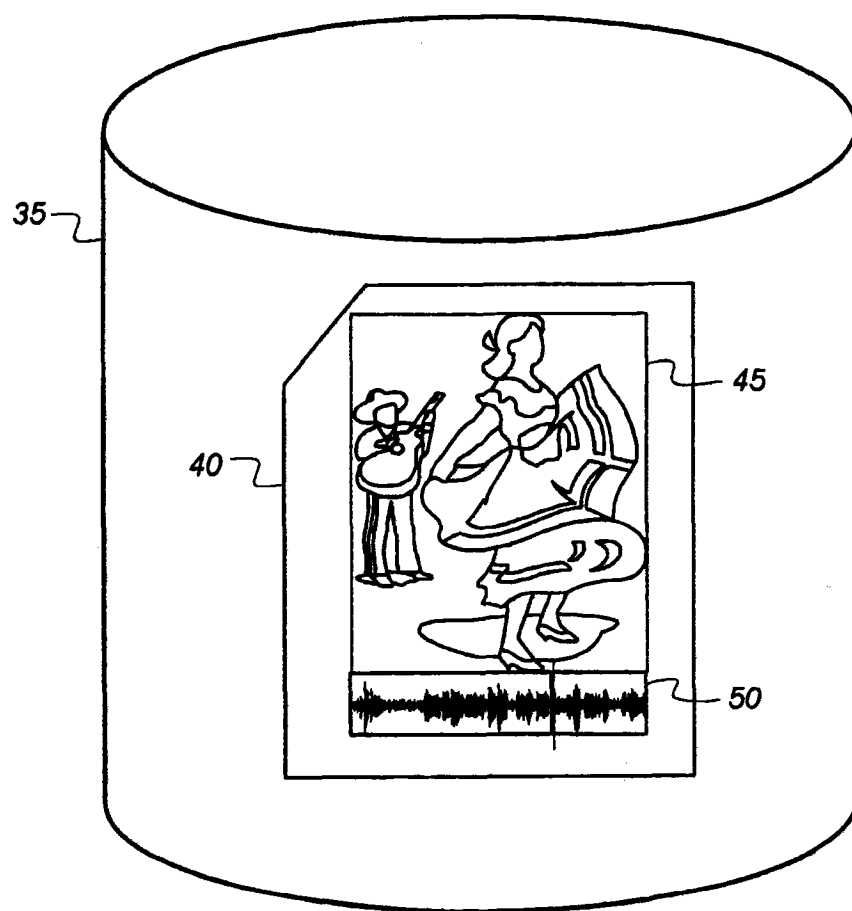


图 1B

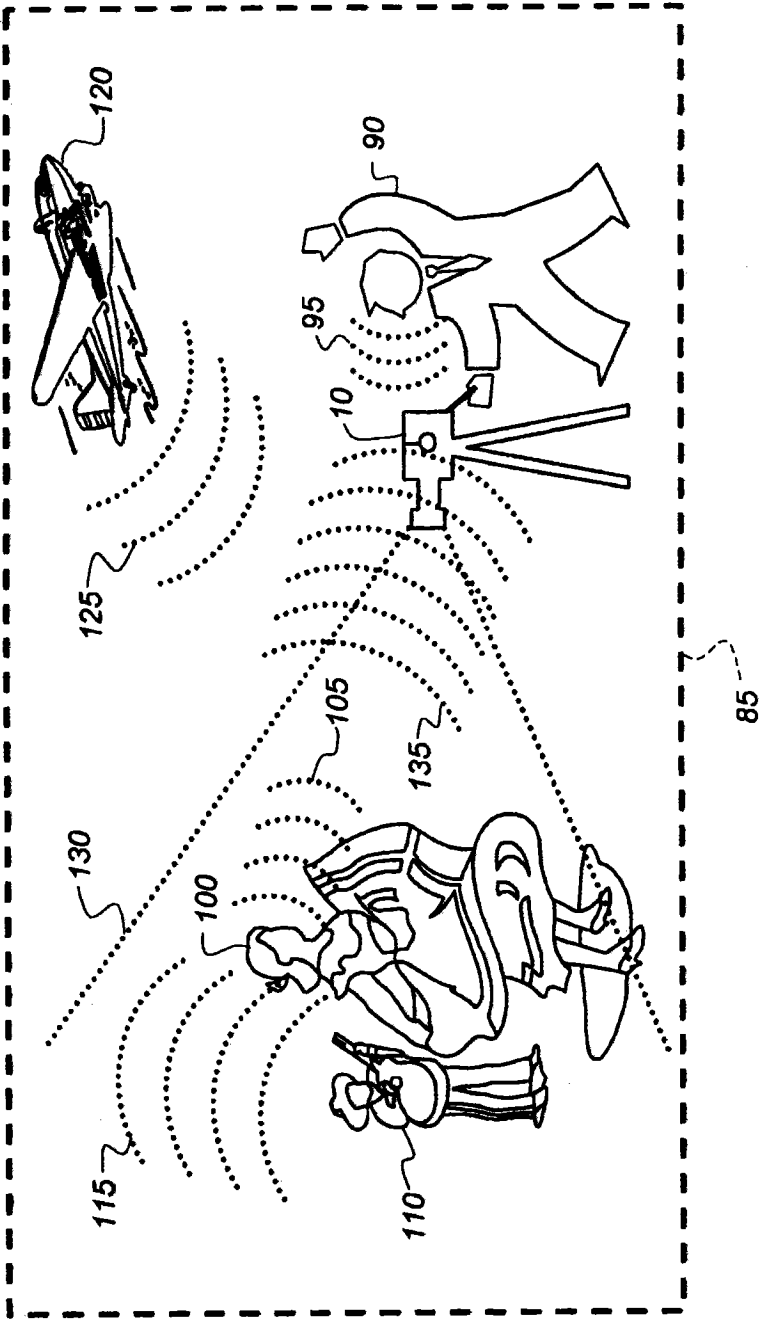


图 2A

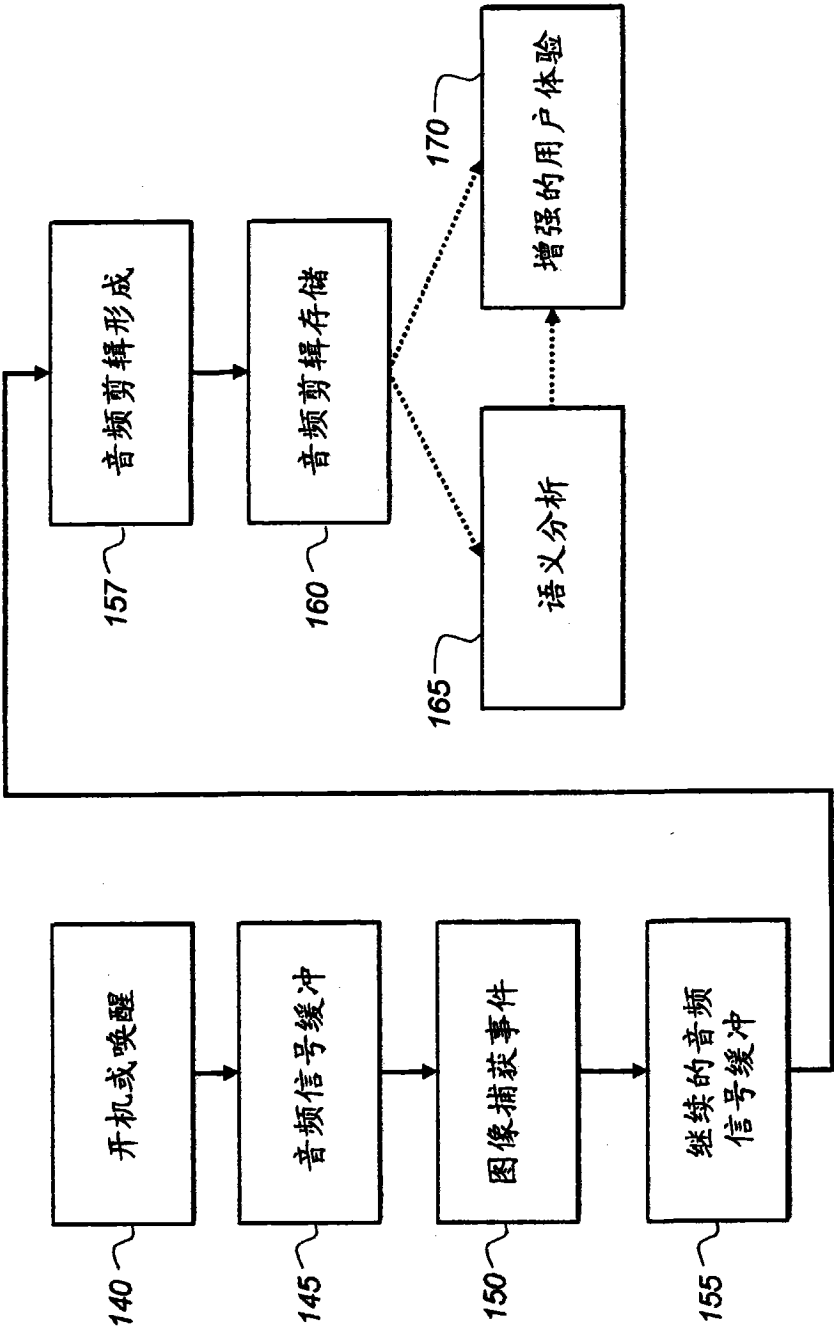


图 2B

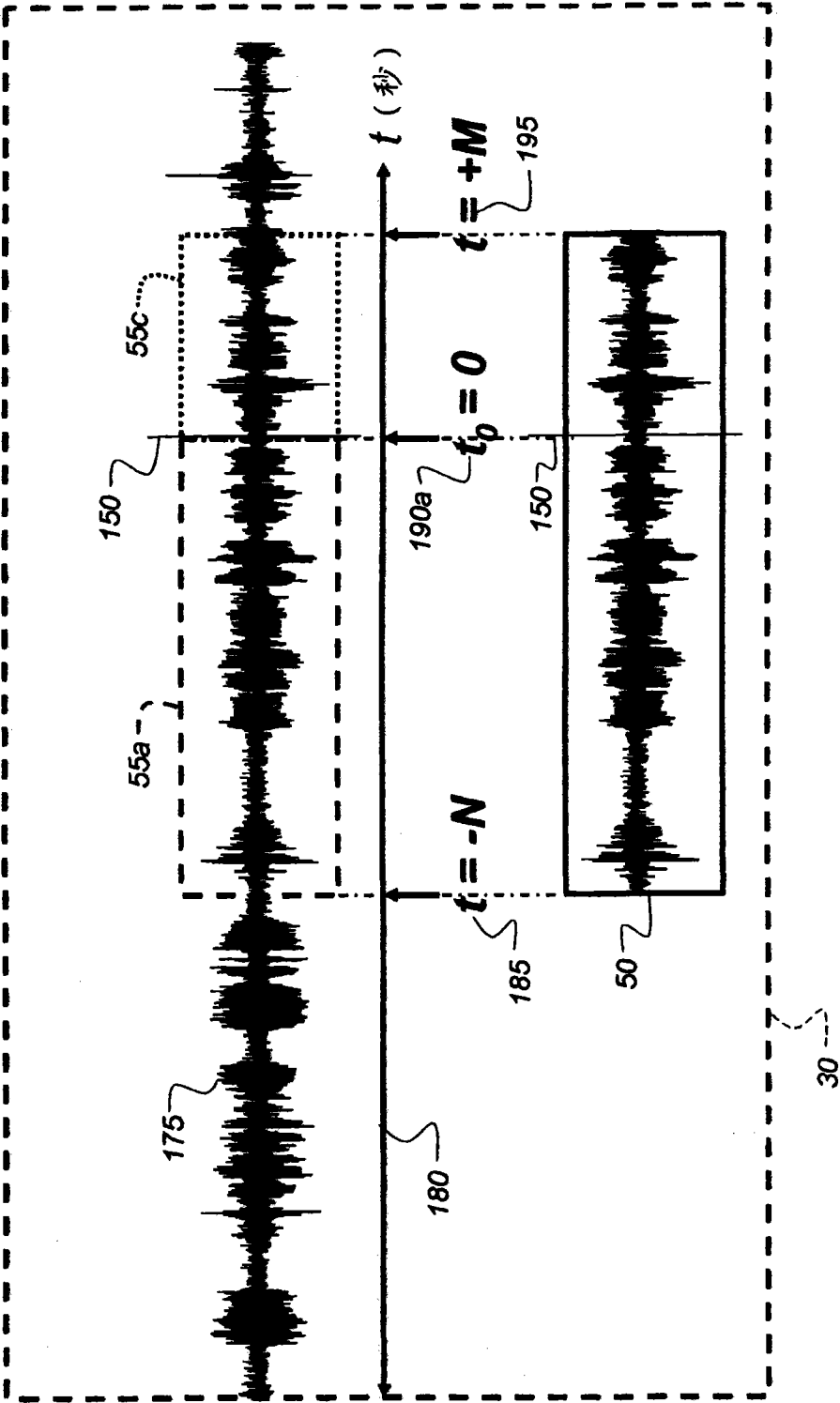


图 3A

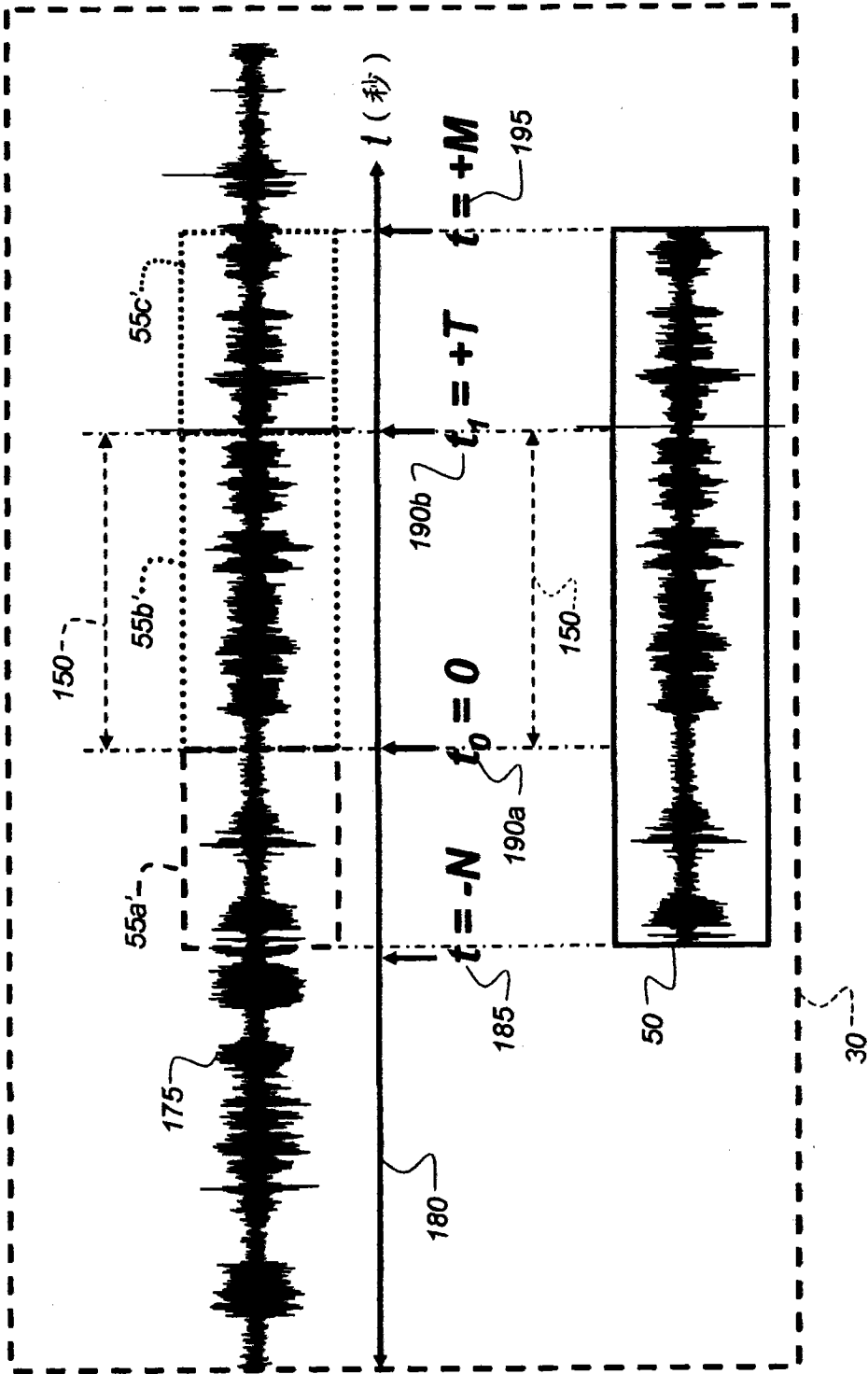


图 3B

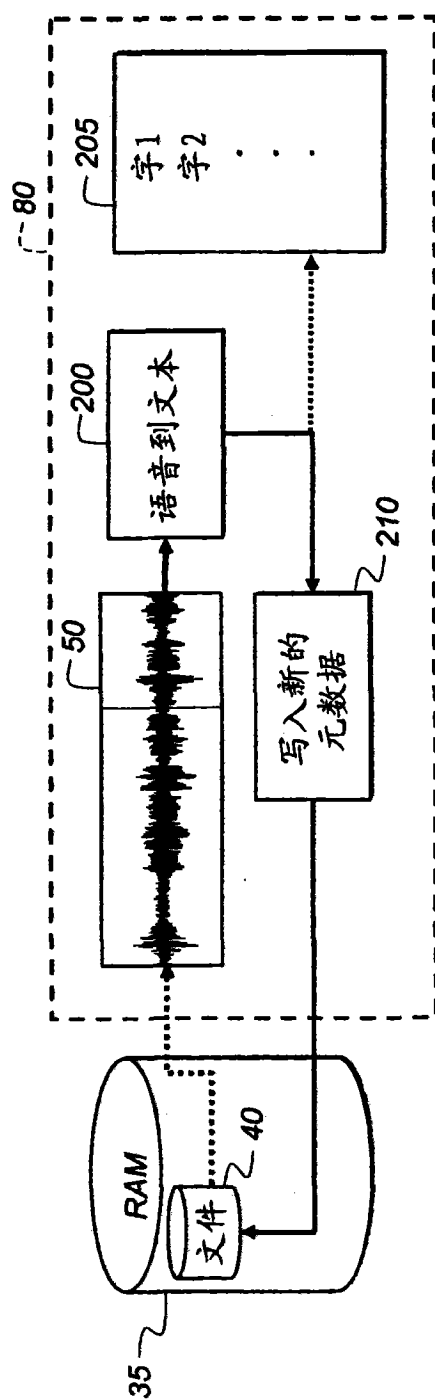


图 4