(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 101640042 B
(45)授权公告日 2013.03.13

(21)申请号 200910162220.3

(22)申请日 2009.07.29

(30)优先权数据
2008-194800 2008.07.29 JP

(73)专利权人 佳能株式会社
地址 日本东京都大田区下丸子 3 丁目 30-2

(72)发明人 山本宽树

(74)专利代理机构 北京魏启学律师事务所 11398
代理人 魏启学

(51)Int.Cl.
G10L 15/26(2006.01)
G03B 17/00(2006.01)
H04N 5/232(2006.01)

(56)对比文件
CN 1506741 A,2004.06.23, 说明书第 6 页第 12-32 行,图 3.
JP 特开平 11-194392 A,1999.07.21, 权利要求 4,说明书第 [0018]-[0021] 段.
JP 特开 2002-354335 A,2002.12.06, 全文.
JP 特开 2006-184589 A,2006.07.13, 全文.
CN 1506741 A,2004.06.23, 说明书第 6 页第 12-32 行,图 3.
US 2005/0102133 A1,2005.05.12, 全文.
CN 1841187 A,2006.10.04, 全文.
CN 1890951 A,2007.01.03, 全文.

审查员 康丹丹

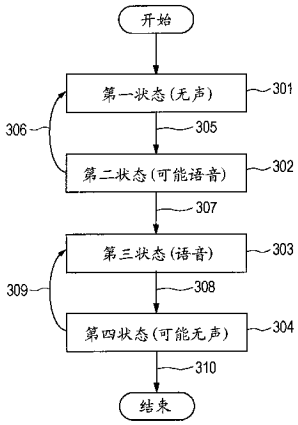
权利要求书 2 页 说明书 26 页 附图 17 页

(54)发明名称

信息处理方法和信息处理设备

(57)摘要

本发明涉及一种信息处理方法和信息处理设备。该信息处理方法包括:检测满足预先设置的标准的第一声音的开始,并且响应于检测到所述第一声音的开始,获得图像数据;或者检测所述第一声音的结束,并且响应于检测到所述第一声音的结束,获得图像数据;将获得的所述图像数据存储在存储器中;以及根据所述第一声音的内容,判断所述图像数据是否是要存储的数据。



CN 101640042 B

1. 一种信息处理方法，包括：

检测满足预先设置的标准的第一声音的开始，并且响应于检测到所述第一声音的开始，获得图像数据；或者检测所述第一声音的结束，并且响应于检测到所述第一声音的结束，获得图像数据；

将获得的所述图像数据存储在存储器中；

获得所述第一声音的语音识别结果；以及

根据所述第一声音的语音识别结果，判断所述图像数据是否是要存储的数据，

其中，如果所述第一声音的语音识别结果是作为用于在检测到所述第一声音的开始时获得图像数据的命令的单词，则将响应于检测到所述第一声音的开始而获得的图像数据存储在存储介质上；如果所述第一声音的语音识别结果是作为用于在检测到所述第一声音的结束时获得图像数据的命令的单词，则将响应于检测到所述第一声音的结束而获得的图像数据存储在所述存储介质上。

2. 根据权利要求 1 所述的信息处理方法，其特征在于，还包括：

在将所述图像数据存储在所述存储介质上之后，或者在所述第一声音的语音识别结果是除作为用于获得图像数据的命令的单词以外的单词的情况下，从所述存储器删除所存储的图像数据。

3. 根据权利要求 1 所述的信息处理方法，其特征在于，

在检测到的所述第一声音的开始的时刻或在检测到的所述第一声音的结束的时刻，执行所述图像数据的获得。

4. 根据权利要求 3 所述的信息处理方法，其特征在于，还包括：

在检测到所述第一声音的开始时，获得图像数据，并且在检测到的所述第一声音的开始的时刻之后所述第一声音没有持续预先设置的时间段的情况下，从所述存储器删除获得的所述图像数据；

检测满足所述预先设置的标准的第二声音的开始；以及

响应于检测到所述第二声音的开始，再次获得图像数据作为第一图像数据。

5. 根据权利要求 3 所述的信息处理方法，其特征在于，还包括：

在检测到所述第一声音的结束时，获得图像数据，并且在检测到的所述第一声音的结束的时刻之后的预先设置的时间段内存在满足所述预先设置的标准的第二声音的情况下，从所述存储器删除获得的所述图像数据；

检测所述第二声音的结束；以及

响应于检测到所述第二声音的结束，获得图像数据作为第二图像数据。

6. 根据权利要求 1 所述的信息处理方法，其特征在于，

在从检测到的所述第一声音的开始的时刻起过去了预先设置的延迟时间段时，或者在从检测到的所述第一声音的结束的时刻起过去了预先设置的延迟时间段时，执行所述图像数据的获得。

7. 根据权利要求 1 所述的信息处理方法，其特征在于，

所述预先设置的标准为音量大于一定水平。

8. 一种信息处理设备，包括：

第一检测单元，用于检测满足预先设置的标准的声音的开始；

第一获得单元，用于响应于检测到所述声音的开始，获得第一图像数据；

第一存储控制单元，用于将所述第一图像数据存储在存储器中；

第二检测单元，用于检测所述声音的结束；

第二获得单元，用于响应于检测到所述声音的结束，获得第二图像数据；

第二存储控制单元，用于将所述第二图像数据存储在所述存储器中；以及

确定单元，用于根据所述声音的内容，将所述第一图像数据和所述第二图像数据中的一个确定为是要存储的数据，并且将所述第一图像数据和所述第二图像数据中的另一个确定为是要删除的数据，

其中，如果所述声音的内容是作为用于在检测到所述声音的开始时获得图像数据的命令的单词，则将响应于检测到所述声音的开始而获得的第一图像数据存储在存储介质上；如果所述声音的内容是作为用于在检测到所述声音的结束时获得图像数据的命令的单词，则将响应于检测到所述声音的结束而获得的第二图像数据存储在所述存储介质上。

9. 一种信息处理方法，包括：

检测满足预先设置的标准的声音的开始；

响应于检测到所述声音的开始，获得第一图像数据；

存储所述第一图像数据；

检测所述声音的结束；

响应于检测到所述声音的结束，获得第二图像数据；

存储所述第二图像数据；以及

根据所述声音的内容，将所述第一图像数据和所述第二图像数据中的一个确定为是要存储的数据，并且将所述第一图像数据和所述第二图像数据中的另一个确定为是要删除的数据，

其中，如果所述声音的内容是作为用于在检测到所述声音的开始时获得图像数据的命令的单词，则将响应于检测到所述声音的开始而获得的第一图像数据存储在存储介质上；如果所述声音的内容是作为用于在检测到所述声音的结束时获得图像数据的命令的单词，则将响应于检测到所述声音的结束而获得的第二图像数据存储在所述存储介质上。

10. 一种信息处理设备，包括：

摄像单元，用于响应于声音的输入来拍摄图像，其中，所述图像是要存储的图像的候选图像；

存储控制单元，用于将拍摄到的所述图像存储在存储器中；

识别结果处理单元，用于获得所述声音的语音识别结果，以及

确定单元，用于根据所述声音的语音识别结果，从存储在所述存储器中的图像中确定作为要存储的图像的图像，

其中，如果所述声音的语音识别结果是作为用于在检测到所述声音的开始时拍摄图像的命令的单词，则将响应于检测到所述声音的开始而拍摄到的图像存储在存储介质上；如果所述声音的语音识别结果是作为用于在检测到所述声音的结束时拍摄图像的命令的单词，则将响应于检测到所述声音的结束而拍摄到的图像存储在所述存储介质上。

信息处理方法和信息处理设备

技术领域

[0001] 本发明涉及一种用于响应于声音开始拍摄图像的技术。

背景技术

[0002] 已知一种具有在检测到大于一定水平的音量时执行图像拍摄的功能（以下称为音量检测快门功能）的照相机（日本特开平 11-194392 号公报）。利用该功能使得能够在发音时拍摄图像。

[0003] 此外，已知一种具有在识别出用于拍摄图像的语音命令时执行图像拍摄的功能（以下称为语音识别快门功能）的照相机（日本特开 2006-184589 号公报）。利用该功能使得能够在用户期望拍摄图像且发音时拍摄图像。这里，当利用具有语音识别快门功能的照相机拍摄图像时，即使用户发出了用于拍摄图像的语音命令，在用户完全发出用于拍摄图像的语音命令之前也不执行照相机的摄像操作。因此，可能错失期望拍摄图像的时机。

[0004] 相反，当利用具有现有的音量检测快门功能的照相机拍摄图像时，可以响应于发出语音的时刻执行摄像操作。然而，在这种情况下，即使当检测到例如除期望的语音以外的大的噪声等声音时，也执行摄像操作。因此，存在可能存储不期望的图像的情况。

[0005] 例如，通过使照相机进行下面的处理可以解决上述问题：根据用户说出的单词“Shoot”（拍摄）在用户期望的时刻拍摄图像的处理和根据语音命令“Delete”（删除）删除拍摄到的图像的处理。然而，输入两个不同的语音命令导致效率不高。

[0006] 根据现有的例子做出了本发明。根据本发明，按照单个语音命令，高效地对在反映输入了特定声音的时刻的时刻拍摄的且作为用户期望的图像的图像进行存储。

发明内容

[0007] 为了高效地存储这种图像，例如，根据本发明的数据转换设备具有下面的结构。

[0008] 根据本发明的实施例，一种信息处理方法包括：检测满足预先设置的标准的第一声音的开始，并且响应于检测到所述第一声音的开始，获得图像数据，或者检测所述第一声音的结束，并且响应于检测到所述第一声音的结束，获得图像数据；将获得的所述图像数据存储在存储器中；以及根据所述第一声音的内容，判断所述图像数据是否是要存储的数据。

[0009] 根据本发明的另一实施例，一种信息处理设备包括：第一检测单元，用于检测满足预先设置的标准的声音的开始；第一获得单元，用于响应于检测到所述声音的开始，获得第一图像数据；第一存储控制单元，用于将所述第一图像数据存储在存储器中；第二检测单元，用于检测所述声音的结束；第二获得单元，用于响应于检测到所述声音的结束，获得第二图像数据；第二存储控制单元，用于将所述第二图像数据存储在所述存储器中；以及确定单元，用于根据所述声音的内容，将所述第一图像数据和所述第二图像数据中的一个确定为是要存储的数据，并且将所述第一图像数据和所述第二图像数据中的另一个确定为是要删除的数据。

[0010] 根据本发明的另一实施例，一种信息处理方法包括：检测满足预先设置的标准的

声音的开始；响应于检测到所述声音的开始，获得第一图像数据；存储所述第一图像数据；检测所述声音的结束；响应于检测到所述声音的结束，获得第二图像数据；存储所述第二图像数据；以及根据所述声音的内容，将所述第一图像数据和所述第二图像数据中的一个确定为是要存储的数据，并且将所述第一图像数据和所述第二图像数据中的另一个确定为是要删除的数据。

[0011] 根据本发明的另一实施例，一种信息处理设备包括：摄像单元，用于响应于声音的输入来拍摄图像，其中，所述图像是要存储的图像的候选图像；存储控制单元，用于将拍摄到的所述图像存储在存储器中；以及确定单元，用于根据所述声音的内容，从存储在所述存储器中的图像中确定作为要存储的图像的图像。

[0012] 通过以下参考附图对典型实施例的说明，本发明的其它特征将变得明显。

附图说明

[0013] 图 1 是示出根据本发明第一实施例的信息处理设备的结构的例子的功能框图；

[0014] 图 2A 和 2B 是本发明第一实施例所使用的数字照相机的外视图；

[0015] 图 3 是示出通过语音检测单元所确定的状态的例子的图；

[0016] 图 4 是示出语音检测单元的操作的例子的概略图；

[0017] 图 5 是由语音检测单元进行的处理操作的流程图；

[0018] 图 6 是示出在通过语音命令拍摄图像时由数字照相机进行的处理的例子的第一流程图；

[0019] 图 7 是示出在通过语音命令拍摄图像时由数字照相机进行的处理的例子的第二流程图；

[0020] 图 8 是示出在通过语音命令拍摄图像时由数字照相机进行的处理的例子的第三流程图；

[0021] 图 9 是示出本发明第一实施例所利用的语音识别语法的例子的图；

[0022] 图 10 是示出识别结果控制表的例子的图；

[0023] 图 11 是示出利用根据本发明第一实施例的数字照相机通过语音命令“Shoot”来拍摄图像的情况下的操作的图；

[0024] 图 12 是示出利用根据本发明第一实施例的数字照相机通过语音命令“Cheese”（笑一笑）来拍摄图像的情况下的操作的图；

[0025] 图 13 是仅在检测到的发音开始的时刻拍摄图像的情况下的流程图；

[0026] 图 14A 和 14B 是示出由信息处理设备进行的处理操作的例子的第一流程图；

[0027] 图 15 是示出由信息处理设备进行的处理操作的例子的第二流程图；

[0028] 图 16 是示出根据本发明第二实施例的信息处理设备的结构的例子的功能框图。

具体实施方式

[0029] 下面，参考附图说明根据本发明的实施例。

[0030] 图 1 是示出作为根据第一实施例的信息处理设备的结构的例子的数字照相机的功能框图。

[0031] 在图 1 中，数字照相机 200 包括控制单元 101、操作单元 102、摄像单元 103、存储器

（用于存储图像）110 和存储介质（用于存储图像）111。

[0032] 此外，数字照相机 200 包括麦克风 112、存储器（用于存储语音识别数据）113、存储器（用于存储识别结果控制表）114 和显示器 115。下面，将对上述单元进行具体说明。

[0033] 控制单元 101 对操作单元 102、摄像单元 103、存储器（用于存储图像）110、存储介质（用于存储图像）111、麦克风 112、存储器（用于存储语音识别数据）113、存储器（用于存储识别结果控制表）114 和显示器 115 的操作进行控制。

[0034] 这里，将在后面说明由控制单元 101 进行的处理。

[0035] 此外，控制单元 101 包括中央处理单元 (CPU)、只读存储器 (ROM) 和随机存取存储器 (RAM) 等。

[0036] 此外，控制单元 101 包括作为软件模块的操作控制单元 122、摄像控制单元 123、图像存储控制单元 104、语音输入单元 105、语音检测单元 106、语音识别单元 107、识别结果处理单元 108 和显示控制单元 109。

[0037] 操作控制单元 122 是用于检测用户对操作单元 102 进行的操作的单元。

[0038] 摄像控制单元 123 是用于使摄像单元 103 执行摄像操作的单元。

[0039] 图像存储控制单元 104 控制将数据写入存储器（用于存储图像）110 和存储介质（用于存储图像）111，并且控制从存储器（用于存储图像）110 和存储介质（用于存储图像）111 读取数据和删除数据等。

[0040] 语音输入单元 105 是用于将通过麦克风 112 输入的声音转换成数字音频信号并输出该数字音频信号的单元。

[0041] 语音检测单元 106 以一帧为单位，连续处理从语音输入单元 105 提供的数字音频信号，并且检测满足标准的对象声音。

[0042] 也就是说，语音检测单元 106 从所接收到的音频信号中识别与对象声音相对应的时间段。具体地，语音检测单元 106 以一帧为单位，连续处理音频信号，并且将从检测到满足开始条件的音频信号起直到检测到满足结束条件的音频信号为止的音频信号的区间识别为对象声音。这里，对象声音为例如发音、鼓掌声或口哨声。

[0043] 以下，将说明对象声音是发音的情况。另外，“检测到发音开始”意为检测到满足开始条件的音频信号，并且“检测到发音结束”意为检测到满足结束条件的音频信号。

[0044] 这里，发音期间包括在用户发音的期间（时间段）内，并且是从检测到发音开始时起直到检测到发音结束时为止的时间段。

[0045] 这里，帧是用于将随着时间改变的音频信号分割成各自具有固定时间长度（例如，25.6 毫秒）的区间的处理单位。这里，可以使用相应数量的帧表示时间。

[0046] 语音识别单元 107 包括作为软件模块的声学分析单元和搜索单元，并且识别包括在用户发音的时间段中的命令（称之为语音命令）。

[0047] 这里，命令是可以由语音识别单元 107 识别的声音的组合。该命令的例子有“Shoot”。

[0048] 声学分析单元以一帧为单位分析音频信号，并且输出例如梅尔频率倒谱系数 (Mel frequency cepstrum coefficient，MFCC) 等特征数据。

[0049] 搜索单元使用维特比 (Viterbi) 算法等现有算法进行搜索处理，并且输出预定数量的命令和相应的识别得分，作为识别结果。

[0050] 此外，在执行搜索处理时，搜索单元使用包括在存储器（用于存储语音识别数据）113 中的声学模型和语言模型。

[0051] 这里，将在后面具体说明声学模型和语言模型。

[0052] 这里，识别得分可以是表示声学相似度的现有声学得分、从语言模型获得的现有语言得分、或加权声学得分和加权语言得分的总和。此外，识别得分可以是表示识别结果的置信度的现有置信度得分。

[0053] 这里，可以通过使用不同得分或多种得分对各种声音执行适当的搜索处理。

[0054] 识别结果处理单元 108 获得由语音识别单元 107 输出的识别结果，并且通过参考存储在存储器（用于存储识别结果控制表）114 中的识别结果控制表，确定与包括在识别结果中的命令相对应的控制。

[0055] 这里，将在后面说明第一实施例中所使用的识别结果控制表的例子。

[0056] 显示控制单元 109 控制显示在显示器 115 上的显示内容。

[0057] 操作单元 102 是用户手动操作数字照相机 200 的单元。

[0058] 这里，操作单元 102 包括按钮或开关等。

[0059] 摄像单元 103 生成通过镜头所形成的图像的摄像信号，并且对所生成的摄像信号进行模拟 - 数字 (A/D) 转换等图像处理。

[0060] 这里，摄像单元 103 包括镜头和摄像传感器等。

[0061] 存储器（用于存储图像）110 临时存储由摄像单元 103 所拍摄的图像的图像数据。这里，存储器（用于存储图像）110 是 RAM 等。

[0062] 在数字照相机 200 所进行的处理结束时，存储介质（用于存储图像）111 存储由摄像单元 103 所拍摄的图像的图像数据。这里，存储介质（用于存储图像）111 是非易失性存储器。

[0063] 存储器（用于存储图像）110 用作第一存储器，并且存储介质（用于存储图像）111 用作第二存储器。

[0064] 麦克风 112 接收输入的用户语音，并将输入的语音数据输出至语音输入单元 105。

[0065] 这里，麦克风 112 是现有的单声道麦克风或现有的立体声麦克风等。

[0066] 存储器（用于存储语音识别数据）113 存储用以执行语音识别的数据、例如隐马尔可夫模型 (hidden Markov model，HMM) 等现有声学模型和 N-gram 或随机语法等现有语言模型。

[0067] 这里，N-gram 是通过使用 N 个单词链概率来计算语言概率的语言模型。

[0068] 此外，可以使用写入了能在语音识别中识别出的特定单词和单词之间的连接规则的语音识别语法，作为语言模型。这里，将在后面说明第一实施例所使用的语音识别语法的例子。

[0069] 此外，存储器（用于存储语音识别数据）113 是非易失性存储器等。

[0070] 存储器（用于存储识别结果控制表）114 存储识别结果控制表。此外，存储器（用于存储识别结果控制表）114 是非易失性存储器。

[0071] 这里，将在后面说明第一实施例所使用的识别结果控制表的例子。

[0072] 这里，这种非易失性存储器可以是现有的硬盘、现有的紧凑型闪存卡或安全数字 (Secure Digital，SD) 卡等。

[0073] 此外，这种非易失性存储器还可以是紧凑型光盘 (CD) 或数字多功能光盘 (DVD)。

[0074] 此外，这种非易失性存储器还可以是能通过局域网 (LAN) 适配器或通用串行总线 (USB) 适配器等接口连接至信息处理设备的外部存储介质。

[0075] 显示器 115 显示由摄像单元 103 所拍摄的图像以及存储在信息处理设备和存储介质 (用于存储图像)111 等中的图像。

[0076] 此外，显示器 115 为例如液晶显示器 (LCD) 或有机电致发光 (electroluminescence，EL) 显示器等。

[0077] 图 2A 和 2B 是根据本发明第一实施例的数字照相机的外视图。这里，图 2A 是数字照相机 200 的正面的外视图。图 2B 是数字照相机 200 的背面的外视图。

[0078] 这里，通过相同的附图标记表示与图 1 所示的组件相同的组件，并且将省略对其的说明。

[0079] 在图 2A 和 2B 中，数字照相机 200 包括快门按钮 201、语音快门 on-off (打开 - 关闭) 开关 202、模式拨盘 203、四向选择按钮 204、确定按钮 205、电源按钮 206 和记录按钮 207。这些组件对应于图 1 所示的操作单元 102。

[0080] 下面，将说明数字照相机 200 的各种单元。

[0081] 快门按钮 201 是用于发出拍摄图像的命令的快门按钮。

[0082] 语音快门 on-off 开关 202 是对于是否使用用于根据语音命令执行摄像操作的功能进行切换的开关。

[0083] 模式拨盘 203 是用于通过旋转将数字照相机 200 的操作模式切换成现有的拍摄模式和现有的重放模式等中的一个的模式拨盘。

[0084] 四向选择按钮 204 是用于输入垂直或水平移动某物的命令的四向选择按钮。

[0085] 确定按钮 205 是用于执行特定操作的按钮。

[0086] 电源按钮 206 是用于打开 / 关闭数字照相机 200 的电源的电源按钮。

[0087] 记录按钮 207 是用于手动输入输入语音的开始和结束的按钮。

[0088] 接着，将具体说明语音检测单元 106 的功能。

[0089] 语音检测单元 106 检测满足第一预定标准 (开始条件) 的声音。当语音检测单元 106 检测到满足第一预定标准 (开始条件) 的声音时，语音检测单元 106 进行用于检测满足第二预定标准的声音的检测操作。

[0090] 在从检测到满足第一预定标准 (开始条件) 的声音时开始过去了预先设置的时间之后，语音检测单元 106 判断为检测到的声音是满足第二预定标准的声音。

[0091] 语音检测单元 106 根据输入的音频信号的变化，判断为检测到的声音不是满足第一预定标准 (开始条件) 的声音。也就是说，语音检测单元 106 取消用于检测满足第一预定标准的声音的检测操作。

[0092] 类似地，语音检测单元 106 检测不满足第二预定标准 (结束条件) 的声音。当语音检测单元 106 检测到不满足第二预定标准 (结束条件) 的声音时，语音检测单元 106 进行用于检测不满足第二预定标准的声音的检测操作。

[0093] 在从检测到不满足第二预定标准 (结束条件) 的声音时开始过去了预先设置的时间之后，语音检测单元 106 判断为检测到的声音不是满足第二预定标准的声音。

[0094] 语音检测单元 106 根据输入的音频信号的变化，判断为检测到的声音是满足第二

预定标准（结束条件）的声音。也就是说，语音检测单元 106 取消用于检测不满足第二预定标准的声音的检测操作。

[0095] 图 3 是示出由语音检测单元 106 所确定的检测状态的例子的图。

[0096] 语音检测单元 106 根据音频信号的检测状况，从所处的四种状态中的一种状态改变成另一状态。

[0097] 第一状态 301 是紧挨在开始输入声音之后进入的状态，即没有检测到发音的状态（以下将该状态称为无声 (SILENCE)）。

[0098] 第二状态 302 是进行了用于检测满足预定标准的发音的开始的检测操作但是未设置发音开始的状态（以下将该状态称为可能语音 (POSSIBLE SPEECH)）。

[0099] 第三状态 303 是设置了满足预定标准的发音的开始的状态（以下将该状态称为语音 (SPEECH)）。

[0100] 第四状态 304 是进行了用于检测满足预定标准的发音的结束的检测操作的状态但未设置发音结束的状态（以下将该状态称为可能无声 (POSSIBLE SILENCE)）。

[0101] 这里，在第一实施例中说明了将发音的检测状况（以下简称为“声音检测状况”）分成四种状态的例子。然而，即使组合第二状态 302 和第四状态 304，将声音检测状况分成三种状态，并且判断为声音检测状况是三种状态中的一种，也获得与第一实施例的效果相同的效果。

[0102] 在第一状态 301 下，如果进行了用于检测发音开始的检测操作（如果进行了用于检测从麦克风 112 输入的且满足预定标准的发音的输入开始的检测操作），则检测状态改变成第二状态 302。以附图标记 305 表示该操作。

[0103] 在第二状态 302 下，如果取消用于检测发音开始的检测操作，则检测状态改变成第一状态 301。以附图标记 306 表示该操作。

[0104] 此外，在第二状态 302 下，如果设置了发音开始，则检测状态改变成第三状态 303。以附图标记 307 表示该操作。

[0105] 在第三状态 303 下，如果进行了用于检测发音结束的检测操作（如果进行了结束从麦克风 112 输入的且满足预定标准的发音的输入），则检测状态改变成第四状态 304。以附图标记 308 表示该操作。

[0106] 在第四状态 304 下，如果取消用于检测发音结束的检测操作，则检测状态改变成第三状态 303。以附图标记 309 表示该操作。

[0107] 此外，在第四状态 304 下，如果设置了满足预定标准的发音的结束，则结束用于检测发音的检测操作。以附图标记 310 表示该操作。

[0108] 当在第四状态 304 下设置了发音结束时，结束用于检测发音的检测操作。因此，在进行后面将说明的语音识别处理时，可以抑制用于进行语音检测处理的计算量和功耗等。

[0109] 这里，在第四状态 304 下设置了发音结束的情况下，检测状态可以改变成第一状态 301。

[0110] 检测状态从第四状态 304 改变成第一状态 301 使得能够连续进行用于检测下一发音的检测操作。

[0111] 图 4 是示出由语音检测单元 106 进行的处理的例子的概略图。

[0112] 图 4 示出用户说出单词“Shoot”的情况。

[0113] 这里，“Shoot”是用于开始拍摄图像的命令的例子。下面将说明命令的内容。

[0114] 在图 4 中，以附图标记 420 表示音频信号。

[0115] 此外，以附图标记 421 表示音频信号 420 的区间。区间 421 中的音频信号不是用户发音的音频信号，而是检测到的噪声的音频信号。

[0116] 此外，以附图标记 422 表示音频信号 420 的区间。区间 422 中的音频信号表示由用户说出的“Shoot”的声音。

[0117] 根据第一实施例的语音检测单元 106 进行用于检测发音音量的检测操作，其中，在判断发音是否满足预定标准时使用该音量。

[0118] 这里，如果发音的音量变得大于或等于预定阈值，则进行用于检测发音开始的检测操作，并且如果该音量变得小于预定阈值，则进行用于检测发音结束的检测操作。也就是说，发音满足开始条件的状态意为发音的音量变得大于或等于预定阈值的状态。同时，发音满足结束条件的状态意为发音的音量变得小于预定阈值的状态。

[0119] 在图 4 中，以附图标记 401 表示利用现有方法从音频信号 420 获得的音量 (E(t))。以附图标记 402 表示进行用于检测发音开始的检测操作所使用的阈值 (TH1)。以附图标记 403 表示进行用于检测发音结束的检测操作所使用的阈值 (TH2)。

[0120] 这里，E(t) 表示在时刻 t 开始的帧处的音量。

[0121] 也就是说，如果在第一状态 301 下音量 E(t) ≥ TH1，则进行用于检测发音开始的检测操作，并且如果在第三状态 303 下音量 E(t) < TH2，则进行用于检测发音结束的检测操作。

[0122] 此外，可以使用相同的阈值 (TH1 = TH2) 来进行用于检测发音开始和发音结束的检测操作。

[0123] 此外，如果预定数量的帧满足进行用于检测发音开始的检测操作所使用的条件 (E(t) ≥ TH1)，则设置发音开始。

[0124] 类似地，如果预定数量的帧满足进行用于检测发音结束的检测操作所使用的条件 (E(t) < TH2)，则设置发音结束。

[0125] 在第一实施例中，以 D1（例如，4 个帧）表示用以设置发音开始的帧的数量，并且以 D2（例如，6 个帧）表示用以设置发音结束的帧的数量。

[0126] 因此，如果在检测状态改变成第二状态 302 之后检测到 D1 个帧满足 E(t) ≥ TH1，则设置发音开始，并且检测状态改变成第三状态 303。

[0127] 此外，如果在检测状态改变成第二状态 302 之后且在检测到 D1 个帧之前音量变成 E(t) < TH1，则检测状态改变成第一状态 301。

[0128] 这里，用于将检测状态从第二状态 302 改变成第一状态 301 的处理对应于取消用于检测发音开始的检测操作的处理。

[0129] 类似地，如果在检测状态改变成第四状态 304 之后检测到 D2 个帧满足 E(t) < TH2，则设置发音结束，并且结束语音检测。

[0130] 此外，如果在检测状态改变成第四状态 304 之后且在检测到 D2 个帧之前音量变成 E(t) ≥ TH2，则检测状态改变成第三状态 303。

[0131] 这里，用于将检测状态从第四状态 304 改变成第三状态 303 的处理对应于取消用于检测发音结束的检测操作的处理。

[0132] 这里，作为用以设置发音开始的帧的数量的 D1 通常小于作为用以设置发音结束的帧的数量的 D2；然而，它们可以是相同的 (D1 = D2)。

[0133] 以附图标记 430 表示对于音频信号 420 的语音检测单元 106 的检测状态。

[0134] 第一状态 301 是开始语音输入之后的状态。

[0135] 在音量 401 变得大于或等于阈值 TH1 的时刻 t1 开始的帧处，进行用于检测发音开始的检测操作。以附图标记 404 表示该操作。检测状态改变成第二状态 302。

[0136] 在检测状态已改变成第二状态 302 之后帧的数量变成 D1 之前的时刻 t2 开始的帧处，音量 401 变得小于阈值 TH1。因此，取消用于检测发音开始的检测操作。以附图标记 405 表示该操作。检测状态改变成第一状态 301。

[0137] 然后，在时刻 t3 开始的帧处，音量 401 再次变得大于或等于阈值 TH1。因此，进行用于检测发音开始的检测操作。以附图标记 406 表示该操作。检测状态改变成第二状态 302。

[0138] 在检测状态已改变成第二状态 302 之后音量 401 大于或等于阈值 TH1 的帧的数量变成 D1 的时刻 t4，将发音开始确定为时刻 t3。以附图标记 407 表示该操作。检测状态改变成第三状态 303。

[0139] 在第三状态 303 下，在音量 401 变得小于进行用于检测发音结束的检测操作所使用的阈值 TH2 的时刻 t5 开始的帧处，进行用于检测发音结束的检测操作。以附图标记 408 表示该操作。检测状态改变成第四状态 304。

[0140] 由于在时刻 t6 开始的帧处音量 401 变得大于或等于阈值 TH2，因而取消用于检测发音结束的检测操作。以附图标记 409 表示该操作。检测状态改变成第三状态 303。

[0141] 由于在时刻 t7 开始的帧处音量 401 再次变得小于阈值 TH2，因而进行用于检测发音结束的检测操作。以附图标记 410 表示该操作。检测状态改变成第四状态 304。

[0142] 此后，在检测状态已改变成第四状态 304 之后音量 401 变得小于阈值 TH2 的帧的数量变成 D2 的时刻 t8，将发音结束确定为时刻 t7。以附图标记 411 表示该操作。

[0143] 此外，代替帧的数量，可以根据音量大于或等于阈值的状态和音量小于阈值的状态是否分别保持预定时间段，来设置发音开始和发音结束。

[0144] 也就是说，如果在与用以设置发音开始的帧的数量 D1（例如，4 个帧）相对应的时间段 S1(40 毫秒）内，检测到音量大于或等于阈值 (TH1)，则设置发音开始。

[0145] 类似地，如果在与用以设置发音结束的帧的数量 D2（例如，6 个帧）相对应的时间段 S1(60 毫秒）内，检测到音量小于或等于阈值 (TH2)，则设置发音结束。

[0146] 这里，即使当检测到间歇检测到预定音量的时间段时，也可以使用该时间段来判断是否应该设置发音开始或发音结束。

[0147] 利用这一结构，即使在呼吸的瞬间没有检测到要检测的声音，并且与该瞬间相对应的帧的音量较低，语音检测单元 106 也可以在该瞬间之后不久再次检测到声音的情况下，执行适当的处理。

[0148] 图 5 是由语音检测单元 106 进行的处理操作的流程图。

[0149] 在步骤 S501，当进行用于检测发音开始的检测操作时，初始化帧编号。

[0150] 以下，以一帧为单位进行用于检测语音的检测操作。

[0151] 也就是说，当语音检测单元 106 以一帧为单位进行处理时，语音检测单元 106 以一

帧为单位计算音量。

[0152] 这里，例如通过利用现有方法根据音频信号计算对数幂等的关于信号强度的值来获得音量。

[0153] 这里，例如通过下面的表达式计算短时间段的对数幂。

[0154] $E(t) = \log\{\sum (x(t,i)^2)/N\}(1 \leqslant i \leqslant N)$ 公式 (1)

[0155] 这里，N 表示每帧的音频信号的样本数量，i 表示帧中的音频信号的样本的索引。

[0156] 此外，x(t，i) 表示在时刻 t 开始的帧中的音频信号的第 i 个样本。

[0157] 此外，x(t，i)^2 意为 x(t，i) 的平方。

[0158] 接着，在步骤 S 502，开始第一状态 301 下的处理。

[0159] 接着，在步骤 S 503，判断在时刻 t 开始的帧处的音量 E(t) 是否大于或等于进行用于检测发音开始的检测操作所使用的阈值 TH 1。

[0160] 如果音量 E(t) 大于或等于阈值 TH1（步骤 S503 为“是”），则在步骤 S 505，检测状态改变成第二状态 302。

[0161] 如果音量 E(t) 小于阈值 TH1（步骤 S503 为“否”），则对于下一帧再次执行处理（步骤 S504）。

[0162] 接着，在步骤 S506，将检测状态改变成第二状态 302 的帧设置为发音开始帧 Ts。

[0163] 接着，在步骤 S507，判断音量 E(t) 是否小于阈值 TH1。

[0164] 如果音量 E(t) 小于阈值 TH1（步骤 S507 为“是”），则检测状态改变成第一状态 301。

[0165] 如果音量 E(t) 大于或等于阈值 TH1（步骤 S507 为“否”），则在步骤 S508 继续该处理，在步骤 S508，判断在检测状态已改变成第二状态 302 之后所获得的帧的数量是否小于 D1。

[0166] 如果在检测状态已改变成第二状态 302 之后所获得的帧的数量小于 D1（步骤 S508 为“是”），则对于下一帧再次执行处理（步骤 S509）。

[0167] 如果在检测状态已改变成第二状态 302 之后所获得的帧的数量大于或等于 D1（步骤 S508 为“否”），则在步骤 S510，检测状态改变成第三状态 303。

[0168] 接着，在步骤 S512，判断音量 E(t) 是否小于进行用于检测发音结束的检测操作所使用的阈值 TH2。

[0169] 如果音量 E(t) 小于阈值 TH2（步骤 S512 为“是”），则在步骤 S514，检测状态改变成第四状态 304。

[0170] 如果 E(t) 大于或等于阈值 TH2（步骤 S512 为“否”），则在步骤 S513 进行下一帧的处理。

[0171] 接着，在步骤 S515，将检测状态改变成第四状态 304 的帧设置为发音结束帧 Te。

[0172] 接着，在步骤 S516，判断音量 E(t) 是否大于或等于阈值 TH2。

[0173] 如果音量 E(t) 大于或等于阈值 TH2（步骤 S516 为“是”），则检测状态改变成第三状态 303。

[0174] 如果音量 E(t) 小于阈值 TH2（步骤 S516 为“否”），则在步骤 S517 继续该处理，在步骤 S517，判断在检测状态已改变成第四状态 304 之后所获得的帧的数量是否小于 D2。

[0175] 如果在检测状态已改变成第四状态 304 之后所获得的帧的数量小于 D2（步骤 S517

为“是”),则在步骤 S518 进行下一帧的处理。

[0176] 如果在检测状态已改变成第四状态 304 之后所获得的帧的数量大于或等于 D2(步骤 S517 为“否”),则在步骤 S519 继续该处理,在步骤 S519,判断是否应该结束语音检测。

[0177] 如果应该结束语音检测(步骤 S519 为“是”),则在步骤 S520 终止语音检测。

[0178] 如果不应该结束语音检测(步骤 S519 为“否”),则在要进行下一发音的检测操作的情况下,检测状态改变成第一状态 301。

[0179] 通过进行上述处理,语音检测单元 106 检测从帧 Ts 开始到帧 Te 为止的发音期间。

[0180] 语音识别单元 107 通过处理在由语音检测单元 106 检测到的发音期间(从帧 Ts 到帧 Te)所获得的音频信号,来获得语音识别结果。

[0181] 这里,使用图 5 的流程图,在上述说明中根据音量的变化,检测发音期间;然而,用于检测发音的检测操作不局限于此。

[0182] 此外,在进行语音检测时,可以使用零交叉次数、音高 (pitch)、从语音模型输出的似然比或从非语音模型输出的似然比等的已知特征或者通过组合这些特征所获得的特征。

[0183] 使用这种特征使得即使在例如输入的周围声音响度大的环境下也能够高效地检测发音开始和发音结束。

[0184] 这里,如下所述,设置发音开始和发音结束所使用的条件可以是除关于帧的数量的条件以外的条件。

[0185] 例如,设置预定阈值 TH3,其中预定阈值 TH3 大于进行用于检测发音开始的检测操作所使用的阈值 TH1。在进行用于检测发音开始的检测操作之后,在音量达到预定阈值 TH3 的帧处,可以将发音开始确定为进行用于检测发音开始的检测操作的时刻。

[0186] 此外,为了设置发音结束,设置小于进行用于检测发音结束的检测操作所使用的阈值 TH2 的预定阈值 TH4。在进行用于检测发音结束的检测操作之后,在音量变得小于预定阈值 TH4 的帧处,可以将发音结束确定为进行用于检测发音结束的检测操作的时刻。

[0187] 使用这种条件可以缩短用于设置发音开始和发音结束的时间段。

[0188] 接着,将说明下面的情况:在具有上述结构的数字照相机 200 中,根据语音命令执行摄像操作。

[0189] 下面参考图 3 说明通过语音检测单元 106、摄像控制单元 123 和图像存储控制单元 104 所进行的处理的例子。

[0190] 在图 3 中,如果进行以附图标记 305 所表示的用于检测发音开始的检测操作,则摄像控制单元 123 使得摄像单元 103 执行摄像操作。

[0191] 这里,进行用于检测发音开始的检测操作 (305) 的情况对应于在图 5 的步骤 S503 中判断为“是”的情况。

[0192] 此外,如果进行以附图标记 308 所表示的用于检测发音结束的检测操作,则摄像控制单元 123 使得摄像单元 103 执行摄像操作。

[0193] 这里,进行用于检测发音结束的检测操作 (308) 的情况对应于在图 5 的步骤 S512 中判断为“是”的情况。

[0194] 也就是说,当语音检测处理的内部状态从第一状态 301 改变成第二状态 302 时,或者当语音检测处理的内部状态从第三状态 303 改变成第四状态 304 时,摄像单元 103 拍摄图像。

[0195] 此外，如果取消以附图标记 306 表示的用于检测发音开始的检测操作，或者如果取消以附图标记 309 表示的用于检测发音结束的检测操作，则图像存储控制单元 104 删除拍摄到的图像。

[0196] 这里，取消用于检测发音开始的检测操作 (306) 的情况对应于在图 5 的步骤 S 507 中判断为“是”的情况。

[0197] 此外，取消用于检测发音结束的检测操作 (309) 的情况对应于在图 5 的步骤 S516 中判断为“是”的情况。

[0198] 也就是说，当在图 3 中取消用于检测发音开始的检测操作时，如果进行用于检测发音开始的检测操作 (305)，则图像存储控制单元 104 删除拍摄到的图像。

[0199] 类似地，当取消用于检测发音结束的检测操作时，如果进行用于检测发音结束的检测操作 (308)，则图像存储控制单元 104 删除拍摄到的图像。

[0200] 也就是说，当内部状态从第二状态 302 改变成第一状态 301，或者当内部状态从第四状态 304 改变成第三状态 303 时，删除紧挨在内部状态改变之前所拍摄到的图像。

[0201] 图 9 是示出第一实施例中所使用的语音识别语法的例子的图。

[0202] 在该例子中，语音识别语法 900 包括描述规则的部分 901 和描述可识别命令和发音的部分 902。

[0203] 在描述可识别命令和发音的部分 902 中，描述了单词的 ID903、关于单词的命令 904 和单词的发音 905。部分 902 的每一行具有其中一个单词的 ID 903、关于该单词的命令 904 和该单词的发音 905。

[0204] 这里，在描述规则的部分 901 中，以语音识别单元 107 可读取的程序代码来描述用于识别部分 902 中所描述的 9 个单词的方法。

[0205] “Shoot”、“Go”（拍了）、“Cheese”、“Say Cheese”（笑一下）和“Five Four Three”（五四三）是用于开始下面所述的摄像操作的语音命令。

[0206] “Spot Metering”（点测光）、“Center Metering”（中央重点测光）、“Use a Flash”（启动闪光灯）和“No Flash”（禁用闪光灯）是用于设置拍摄条件的语音命令。

[0207] 在下面的说明中，使用图 9 所示的语音识别语法 900 作为根据第一实施例的数字照相机 200 中的语言模型。

[0208] 这里，在第一实施例中，作为例子说明了语音命令；然而，本发明不局限于这些。例如，代替语音命令，可以使用能被解释为表示语音命令的声音。

[0209] 例如，可以使用笑声或火车经过时发出的声音等。这里，在这种情况下，代替语音识别技术，使用检测声音内容的已知技术。

[0210] 利用这种结构，即使在通过麦克风 112 不仅输入语音而且输入特征声音的情况下，用户也可以获得在与各种特征声音中的一个相对应的时刻所拍摄的图像。

[0211] 识别结果控制表是表格式的数据，在该数据中，描述了与识别结果相对应的用于拍摄图像的处理、用于启动测光的处理和用于启动闪光灯的处理。识别结果处理单元 108 在确定与识别结果相对应的照相机控制时，参考该识别结果控制表。

[0212] 这里，以识别结果处理单元 108 可读取的程序代码的形式，将识别结果控制表存储在存储器（用于存储识别结果控制表）114 中。

[0213] 图 10 是示出识别结果控制表的例子的图。

[0214] 在图 10 中，以附图标记 1000 表示识别结果处理数据。

[0215] 以附图标记 904 表示语音识别所使用的命令，并且描述了数字照相机 200 的以附图标记 904 表示的命令中的相应一个命令的控制内容，其中，以附图标记 1002 表示该控制内容。

[0216] 图 6 ～图 8 是示出在通过语音命令拍摄图像时由数字照相机 200 所进行的处理的例子的流程图。

[0217] 首先，使用图 6 的流程图来说明处理。

[0218] 在步骤 S601，判断是否启动了声音启动功能。

[0219] 如果启动了声音启动功能（步骤 S601 为“是”），则在步骤 S602 继续该处理，在步骤 S602，判断是否按下了记录按钮 207 和是否进行用于开始语音（发音）输入的操作。

[0220] 如果没有启动声音启动功能（步骤 S601 为“否”），则在步骤 S699 中进行除关于声音启动功能的处理以外的处理（即，其它照相机控制）。

[0221] 这里，用户操作包括在操作单元 102 中的语音快门 on-off 开关 202 以在启动和禁用声音启动功能之间进行切换。

[0222] 此外，控制单元 101 判断应该启动还是禁用声音启动功能。

[0223] 如果进行用于开始接收语音的操作（步骤 S602 为“是”），则在步骤 S603，语音输入单元 105 开始用于接收语音的处理，并且语音检测单元 106 开始语音检测处理。

[0224] 如果进行除用于开始接收语音的操作以外的操作（步骤 S602 为“否”），则在步骤 S699 进行除关于声音启动功能的处理以外的处理（即，其它照相机控制）。

[0225] 这里，可以通过除按下记录按钮 207 以外的操作来进行用于开始接收语音的操作。

[0226] 例如，如果半按下快门按钮 201，则设置有自动调焦功能的数字照相机进行调焦。

[0227] 这里，可以与自动调焦功能的操作相关联地开始用于接收语音的处理。也就是说，如果用户半按下快门按钮 201，则可以开始用于接收语音的处理和用于检测语音的处理。

[0228] 利用这种结构，简化了手动操作。因此，用户可以快速地开始用于输入语音的处理。

[0229] 此外，当向语音输入单元 105 输入音频信号时，可以在无需手动开始语音检测的情况下，开始语音检测。

[0230] 利用这种结构，可以快速地开始用于检测语音的处理。此外，即使用户不能手动操作照相机，用户也可以开始语音检测。因此，可以在监视照相机、安全用照相机或置于高处的照相机等中使用这种结构。

[0231] 在步骤 S604，判断语音检测单元 106 是否进行了用于检测发音开始的检测操作。

[0232] 这里，在步骤 S604，根据语音检测单元 106 是否已执行用于将内部状态从第一状态 301 改变成第二状态 302 的处理，判断语音检测单元 106 是否进行了用于检测发音开始的检测操作。

[0233] 如果语音检测单元 106 进行了用于检测发音开始的检测操作（步骤 S604 为“是”），则在步骤 S605，摄像单元 103 执行摄像操作。

[0234] 在步骤 S606，图像存储控制单元 104 将在前一步骤 S 605 中拍摄到的图像的第一图像数据存储在存储器（用于存储图像）110 中。

[0235] 这里，将在步骤 S605 拍摄到的图像，即在语音检测单元 106 进行用于检测发音开始的检测操作时所拍摄的图像称为图像 A。

[0236] 如果语音检测单元 106 没有进行用于检测发音开始的检测操作（步骤 S604 为“否”），则再次判断语音检测单元 106 是否进行了用于检测发音开始的检测操作。

[0237] 在步骤 S607，判断语音检测单元 106 是否应该取消用于检测发音开始的检测操作。

[0238] 这里，在步骤 S607，根据语音检测单元 106 是否已执行用于将内部状态从第二状态 302 改变成第一状态 301 的处理，判断语音检测单元 106 是否应该取消用于检测发音开始的检测操作。

[0239] 如果取消用于检测发音开始的检测操作（步骤 S607 为“是”），则在步骤 S608 继续该处理，图像存储控制单元 104 删除存储在存储器（用于存储图像）110 中的图像 A。

[0240] 如果没有取消用于检测发音开始的检测操作（步骤 S607 为“否”），则在步骤 S609，判断语音检测单元 106 是否设置了发音开始。

[0241] 这里，在步骤 S609，根据语音检测单元 106 是否执行了用于将内部状态从第二状态 302 改变成第三状态 303 的处理，判断是否设置 / 确定发音开始。

[0242] 如果设置 / 确定了发音开始（步骤 S609 为“是”），则在步骤 S610，语音识别单元 107 开始语音识别处理。

[0243] 如果没有设置 / 确定发音开始（步骤 S609 为“否”），则再次判断是否应该取消用于检测发音开始的检测操作。

[0244] 将参考图 7 的流程图说明下面的处理。

[0245] 在步骤 S711，语音检测单元 106 判断是否进行了用于检测发音结束的检测操作。

[0246] 这里，在步骤 S711，根据语音检测单元 106 是否执行了用于将内部状态从第三状态 303 改变成第四状态 304 的处理，判断是否进行了用于检测发音结束的检测操作。

[0247] 如果进行了用于检测发音结束的检测操作（步骤 S711 为“是”），则在步骤 S712，摄像单元 103 拍摄图像。

[0248] 接着，在步骤 S713，图像存储控制单元 104 将在前一步骤 S712 拍摄到的图像的第二图像数据存储在存储器（用于存储图像）110 中。这里，将在步骤 S712 拍摄到的图像，即在语音检测单元 106 进行用于检测发音结束的检测操作时所拍摄的图像称为图像 B。

[0249] 这里，存在这样一种情况 :通常，在说出了“Say Cheese”等后（发出 /z/ 的音后）过去特定时间段（例如，0.5 秒）之后，拍摄图像。

[0250] 考虑到该情况，在第一实施例中，在语音检测单元 106 进行了用于检测“Say Cheese”发音结束的检测操作后过去预定延迟时间之后，摄像单元 103 拍摄图像。

[0251] 利用这种结构，可以增加用户期望的摄像时刻的种类数量。

[0252] 接着，在步骤 S715，语音检测单元 106 判断是否应该取消用于检测发音结束的检测操作。

[0253] 这里，在步骤 S715，根据语音检测单元 106 是否执行了用于将内部状态从第四状态 304 改变成第三状态 303 的处理，判断是否应该取消用于检测发音结束的检测操作。

[0254] 如果取消了用于检测发音结束的检测操作（步骤 S715 为“是”），则在步骤 S714 继续该处理，在步骤 S714，图像存储控制单元 104 删除存储在存储器（用于存储图像）110 中

的图像 B。

[0255] 接着，在步骤 S716，判断语音检测单元 106 是否应该设置 / 确定发音结束。

[0256] 这里，在步骤 S716，根据语音检测单元 106 是否结束了内部状态的改变并且保持内部状态处于第四状态 304，判断是否应该设置 / 确定发音结束。

[0257] 如果设置 / 确定了发音结束（步骤 S716 为“是”），则在步骤 S717，结束由语音输入单元 105 和语音检测单元 106 所进行的处理。

[0258] 如果没有设置 / 确定发音结束（步骤 S716 为“否），则再次判断是否应该取消用于检测发音结束的检测操作。

[0259] 接着，在步骤 S718，在结束语音检测之后，语音识别单元 107 进行语音识别处理，直到处理了在语音检测单元 106 所检测到的发音期间所获得的所有音频信号为止。

[0260] 如果语音识别处理结束（步骤 S718 为“是”），则在步骤 S719，识别结果处理单元 108 获得由语音识别单元 107 所获得的识别结果。

[0261] 将参考图 8 的流程图说明下面的处理。

[0262] 在步骤 S821，识别结果处理单元 108 判断是接收还是丢弃与所获得的识别结果中的识别得分相对应的命令。

[0263] 这里，接收命令意为控制单元 101 判断为进行与识别出的命令相对应的控制。此外，丢弃命令意为控制单元 101 判断为不进行与识别出的命令相对应的控制。

[0264] 如果所获得的识别得分大于或等于预定阈值，并且接收了相应命令（步骤 S821 为“是”），则在步骤 S822，参考识别结果控制表确定数字照相机 200 的控制，其中，该控制对应于包括在识别结果中的命令。

[0265] 如果识别出的命令是作为用于在发音开始时拍摄图像的命令的单词（“Shoot”或“Go”）（步骤 S822 为“是”），则在步骤 S823，图像存储控制单元 104 将图像 A 的图像数据存储在存储介质（用于存储图像）111 上，其中，图像 A 被存储在存储器（用于存储图像）110 中。

[0266] 这里，步骤 S 823 中的处理是根据识别结果处理单元 108 的判断所进行的处理。

[0267] 接着，在步骤 S824，显示控制单元 109 以用户可以检查拍摄到的图像的方式将图像 A 显示在显示器 115 上。

[0268] 如果识别出的命令不是作为用于在发音开始时拍摄图像的命令的单词（“Shoot”或“Go”）（步骤 S822 为“否”），则在步骤 S826，判断识别出的命令是否是作为用于在发音结束时拍摄图像的命令的单词（“Cheese”）。

[0269] 如果识别出的命令是作为用于在发音结束时拍摄图像的命令的单词（“Cheese”）（步骤 S826 为“是”），则在步骤 S827 继续该处理，在步骤 S827，图像存储控制单元 104 将图像 B 的图像数据存储在存储介质（用于存储图像）111 上。

[0270] 这里，步骤 S827 中的处理是根据识别结果处理单元 108 的判断所进行的处理。

[0271] 在步骤 S828，显示控制单元 109 以用户可以检查拍摄到的图像的方式将图像 B 显示在显示器 115 上。

[0272] 如果识别出的命令是除作为用于拍摄图像的命令的单词以外的单词（“Spot Metering”等）（步骤 S826 为“否”），则在步骤 S 829 继续该处理，在步骤 S 829，识别结果处理单元 108 以进行除用于拍摄图像的控制以外的控制的方式，通过参考识别结果控制

表，控制数字照相机 200。

[0273] 在步骤 S825，图像存储控制单元 104 删除存储在存储器（用于存储图像）110 中的所有图像（图像 A 和 B）的图像数据。

[0274] 也就是说，如果没有识别出预定命令并且丢弃了识别结果，则摄像单元 103 删除拍摄到的图像。

[0275] 该处理丢弃与周围噪声有关的识别结果、识别对象以外的单词的发音、以及用户以外的人的语音等不是想要操作照相机的语音，并且自动删除由于错误检测到这种声音而拍摄的图像。

[0276] 这里，在步骤 S821，判断所使用的阈值可以是预先设置的固定值或者是通过将识别得分乘以 r(0 ＜ r) 所获得的值，其中，利用废料模型 (garbage model) 输出识别得分。

[0277] 废料模型是使用包括语音以外的噪声的噪声或多个估计的未知单词（识别对象以外的单词）所生成的声学模型，并且被包括在存储器（用于存储语音识别数据）113 中。

[0278] 这里，在步骤 S822 ～ S829 的处理中，根据识别结果，将在发音开始时所拍摄的图像和在发音结束时所拍摄的图像中的一个确定为是要存储的图像。

[0279] 因此，用户可以根据发音内容，自由改变要存储的图像的摄像时刻。

[0280] 这里，在上述说明中，在步骤 S825 之后处理结束。然而，该过程可以进入步骤 S602 中的处理，以继续进行下一语音的接收。

[0281] 利用这种结构，如果通过半按下快门按钮 201 开始语音接收，则可以通过在半按下快门按钮 201 的同时尽可能多次地输入语音来进行照相机控制。

[0282] 例如，在半按下快门按钮 201 时，“Center Metering”等的发音可以设置拍摄条件，并且可以通过下一发音来拍摄图像。

[0283] 图 11 是示出使用根据第一实施例的数字照相机 200 利用语音命令“Shoot”来拍摄图像的情况下的操作的图。

[0284] 在图 11 中，水平轴 1150 表示时间，并且时间从左向右推移。附图标记 t1 ～ t7 均表示时刻。

[0285] 附图标记 1110 表示由语音输入单元 105 进行了 A/D 转换的音频信号。

[0286] 附图标记 1111 表示用户说出“Shoot”期间的音频信号（音频波形）。

[0287] 附图标记 1120 表示音量。示出了与音频信号 1110 相对应的音量 1120 的变化。

[0288] 附图标记 1121 表示进行用于检测发音开始的检测操作所使用的且由语音检测单元 106 所使用的阈值 (TH1)。附图标记 1122 表示进行用于检测发音结束的检测操作所使用的且由语音检测单元 106 所使用的阈值 (TH2)。

[0289] 附图标记 1130 表示由语音检测单元 106 识别出的状态。可视地示出了状态 1130 的变化。

[0290] 附图标记 1140 表示数字照相机 200 的操作的细节。

[0291] 接着，将沿着从时刻 t1 至时刻 t7 的时间来说明数字照相机 200 的操作。

[0292] 时刻 t1

[0293] 在音量 1120 变得大于或等于阈值 TH1 的时刻 t1 开始的帧处，语音检测单元 106 进行用于检测发音开始的检测操作。该操作对应于用于检测满足上述第一预定标准（开始条件）的声音的处理。

[0294] 这里，语音检测单元106执行用于将检测状态从第一状态301改变成第二状态302的处理，以时刻t1处的附图标记1130表示该处理。

[0295] 在进行了用于检测发音开始的检测操作的时刻，摄像单元103拍摄被摄体的图像(IMG003)。然后，图像存储控制单元104将拍摄到的图像的图像数据存储在存储器（用于存储图像）110中。以附图标记1141表示这些操作。

[0296] 时刻t2

[0297] 在时刻t2开始的且作为从在时刻t1开始的帧算起的第D1个帧的帧处，语音检测单元106将发音开始确定为时刻t1，其中在时刻t1，进行了用于检测发音开始的检测操作。

[0298] 同时，开始由语音识别单元107所进行的语音识别处理。以附图标记1142表示这些操作。

[0299] 这里，语音检测单元106执行用于将检测状态从第二状态302改变成第三状态303的处理，以时刻t2处的附图标记1130表示该处理。

[0300] 时刻t3

[0301] 接着，在音量1120变得小于阈值TH2的时刻t3开始的帧处，语音检测单元106进行用于检测发音结束的检测操作。在该操作中，检测满足上述预定标准（结束条件）的声音。

[0302] 这里，语音检测单元106执行用于将检测状态从第三状态303改变成第四状态304的处理，以时刻t3处的附图标记1130表示该处理。

[0303] 在语音检测单元106进行用于检测发音结束的检测操作的时刻t3，摄像单元103拍摄被摄体的图像(IMG005)。然后，图像存储控制单元104将拍摄到的图像的图像数据存储在存储器（用于存储图像）110中。以附图标记1143表示这些操作。

[0304] 时刻t4

[0305] 如果在时刻t4开始的帧处，音量1120变得大于或等于阈值TH2，则语音检测单元106取消用于检测发音结束的检测操作，其中在时刻t4开始的帧是作为从在时刻t3开始的帧算起的第D2个帧的帧之前的帧，并且在时刻t3，语音检测单元106进行了用于检测发音结束的检测操作。

[0306] 这里，语音检测单元106执行用于将检测状态从第四状态304改变成第三状态303的处理，以时刻t4处的附图标记1130表示该处理。

[0307] 在取消用于检测发音结束的检测操作的时刻t4，图像存储控制单元104从存储器（用于存储图像）110删除在进行用于检测发音结束的检测操作的时刻t3所拍摄的图像IMG005的图像数据。以附图标记1144表示这些操作。

[0308] 时刻t5

[0309] 在时刻t5开始的帧处，音量1120变得小于阈值TH2，因此语音检测单元106进行用于检测发音结束的检测操作。

[0310] 这里，语音检测单元106执行用于将检测状态从第三状态303改变成第四状态304的处理，以时刻t5处的附图标记1130表示该处理。

[0311] 此外，摄像单元103在时刻t5拍摄被摄体的图像(IMG006)，并且图像存储控制单元104将拍摄到的图像的图像数据存储在存储器(用于存储图像)110中。以附图标记1145表示这些操作。

[0312] 时刻 t6

[0313] 在进行用于检测发音结束的检测操作的时刻 t5 开始的帧和在时刻 t6 开始的且作为从在时刻 t5 开始的帧算起的第 D2 个帧的帧之间，音量 1120 未变得大于或等于阈值 TH2。在时刻 t6 开始的帧处，语音检测单元 106 将发音结束确定为时刻 t5。以附图标记 1146 表示该操作。

[0314] 这里，如上所述，语音检测单元 106 可以执行用于将检测状态从第四状态 304 改变成第一状态 301 的处理，或者语音检测单元 106 可以结束用于改变检测状态的处理。

[0315] 时刻 t7

[0316] 此后，在结束由语音识别单元 107 所进行的处理的时刻 t7，识别结果处理单元 108 确定数字照相机 200 的控制方法。以附图标记 1147 表示该操作。

[0317] 这里，如果获得“Shoot”作为识别结果，则参考识别结果控制表，确定与“Shoot”相对应的处理。

[0318] 如图 10 所示，“Shoot”是与在检测到的发音开始的时刻所进行的摄像操作相关的命令。

[0319] 根据识别结果处理单元 108 的判断，图像存储控制单元 104 将在作为检测到的发音开始的时刻的时刻 t1 拍摄到的图像 (IMG003) 的图像数据存储在存储介质 (用于存储图像)111 中。

[0320] 同时，图像存储控制单元 104 从存储器 (用于存储图像)110 删除在发音结束时所拍摄到的图像 (IMG006)，而不存储该图像。

[0321] 图 12 是示出使用根据第一实施例的数字照相机 200 利用语音命令“Cheese”拍摄图像的情况下的操作的图。

[0322] 类似于图 11，附图标记 1250 表示时间，附图标记 1210 表示音频信号，附图标记 1220 表示音量，附图标记 1230 表示由语音检测单元 106 识别出的状态，附图标记 1240 表示数字照相机 200 的操作。

[0323] 附图标记 1211 表示在用户发音之前碰巧输入的噪声。附图标记 1212 表示由用户说出的语音“Cheese”等。

[0324] 附图标记 1221 表示进行用于检测发音期间的检测操作所使用的阈值 (TH1)，其中语音检测单元 106 使用该阈值 TH1。

[0325] 这里，在图 12 中，使用相同的阈值 TH1 来检测发音开始和发音结束。

[0326] 下面，将沿着时间来说明数字照相机 200 的操作。

[0327] 时刻 t1

[0328] 在时刻 t1 开始的帧处，如果语音检测单元 106 进行用于检测发音开始的检测操作，则摄像单元 103 拍摄与在时刻 t1 开始的帧相对应的被摄体的图像 (IMG001)。此外，图像存储控制单元 104 将拍摄到的图像的图像数据临时存储在存储器 (用于存储图像)110 中。以附图标记 1241 表示这些操作。

[0329] 时刻 t2

[0330] 在时刻 t2 开始的且处于作为从进行用于检测发音开始的检测操作的帧算起的第 D1 个帧的帧之前的帧处，音量 1220 变得小于阈值 TH1，因此语音检测单元 106 取消用于检测发音开始的检测操作。

[0331] 这里，图像存储控制单元 104 删除在操作 1241 中拍摄到的图像（IMG001）。以附图标记 1242 表示这些操作。

[0332] 时刻 t3

[0333] 在时刻 t3 开始的帧处，如果语音检测单元 106 再次进行用于检测发音开始的检测操作，则摄像单元 103 拍摄与在时刻 t3 开始的帧相对应的被摄体的图像（IMG003）。此外，图像存储控制单元 104 将拍摄到的图像的图像数据临时存储在存储器（用于存储图像）110 中。以附图标记 1243 表示这些操作。

[0334] 时刻 t4

[0335] 在时刻 t4 开始的帧处，如果语音检测单元 106 将发音开始确定为时刻 t3，则语音识别单元 107 开始语音识别处理。以附图标记 1244 表示这些操作。

[0336] 时刻 t5

[0337] 在时刻 t5 开始的帧处，如果语音检测单元 106 进行用于检测发音结束的检测操作，则摄像单元 103 拍摄与在时刻 t5 开始的帧相对应的被摄体的图像（IMG005）。此外，然后，图像存储控制单元 104 将拍摄到的图像的图像数据临时存储在存储器（用于存储图像）110 中。以附图标记 1245 表示这些操作。

[0338] 时刻 t6

[0339] 在时刻 t6 开始的帧处，语音检测单元 106 将发音结束确定为时刻 t5。以附图标记 1246 表示该操作。

[0340] 时刻 t7

[0341] 在将发音结束确定为时刻 t5 之后，在结束由语音识别单元 107 所进行的语音识别处理的时刻 t7，识别结果处理单元 108 根据识别结果确定照相机控制。以附图标记 1247 表示这些操作。

[0342] 这里，如图 10 所示，“Cheese”是与在检测到的发音结束的时刻所进行的摄像操作相关的命令。

[0343] 因此，图像存储控制单元 104 将在作为检测到的发音结束的时刻的时刻 t5 所拍摄的图像（IMG005）的图像数据存储在存储介质（用于存储图像）111 中。图像存储控制单元 104 删除在作为检测到的发音开始的时刻的时刻 t3 所拍摄的图像（IMG003）的图像数据，而不存储该图像数据。

[0344] 如以上使用图 11 和图 12 所述，如果要使用第一实施例中所述的数字照相机 200 拍摄发音开始时的图像，则仅要说出“Shoot”（或“Go”）。

[0345] 此外，如果要使用第一实施例中所述的数字照相机 200 拍摄发音结束时的图像，则仅需说出“Cheese”。

[0346] 此外，如果要拍摄从发音开始的时刻起过去了特定时间段的时刻的图像，则仅需说出“Five Four Three”，其中该特定时间段对应于说出“Two One Zero”（二一零）的时间段。

[0347] 此外，如果要拍摄从发音结束的时刻起过去了特定时间段（例如，0.5 秒）的时刻的图像，则仅需说出“Say Cheese”。

[0348] 如果说出“Shoot”（或“Go”），则在结束语音识别之前拍摄图像。因此，这适合于拍摄车辆等运动被摄体的图像的情况。

[0349] 此外，如果说出“Cheese”（或“Say Cheese”），则在发音结束之后拍摄图像。因此，这适合于在通知被摄体拍摄时刻之后拍摄图像的情况，如合影或留念照等。

[0350] 此外，如果说出“Five Four Three”，则可以在从发音开始的时刻起过去了特定时间段之后的时刻拍摄图像，其中该特定时间段对应于说出“Two One Zero”的时间段。

[0351] 因此，可以根据拍摄场景，在任意拍摄时刻拍摄图像，并且提高了用户操作的方便性。

[0352] 此外，在拍摄图像之后，用户可以不必删除在不希望的时刻所拍摄的图像。

[0353] 也就是说，如使用图 12 所述，即使在根据当输入语音时碰巧输入的周围噪声而错误地拍摄了图像的情况下，如果不设置语音开始，则自动删除该图像。

[0354] 此外，即使在利用噪声或不想拍摄图像的发音触发了图像拍摄的情况下，如果在图 8 的步骤 S 821 的处理中识别出不想触发图像拍摄的发音，则丢弃该识别结果，并删除错误拍摄的图像。

[0355] 因此，在利用语音命令触发拍摄开始的情况下，第一实施例具有减少由于周围噪声而导致的误操作发生的效果。

[0356] 在第一实施例中，可以在进行用于检测发音开始的检测操作的时刻拍摄图像，或者可以在进行用于检测发音结束的检测操作的时刻拍摄图像。

[0357] 图 13 是仅在检测到的发音开始的时刻拍摄图像的情况下的流程图。

[0358] 图 13 所示的流程图示出了与使用图 6 ～图 8 的流程图所述的处理不同的步骤 S711 和其后步骤中的处理。

[0359] 此外，以相同的附图标记表示与图 7 和图 8 中的处理相同的处理。在下面，将仅说明图 13 与图 7 和图 8 之间的不同。

[0360] 在图 13 所示的流程图中，不进行图 7 的流程图中的以下处理 :用于在进行用于检测发音结束的检测操作的时刻拍摄图像的处理（步骤 S712 和 S713）和用于删除拍摄到的图像的处理（步骤 S714）。

[0361] 此外，在图 13 所示的流程图中，不进行图 8 的流程图中的以下处理 :在识别出作为用于在发音结束时拍摄图像的命令的单词的情况下，由识别结果处理单元 108 所进行的处理（步骤 S826、S827 和 S828）。

[0362] 其它处理与使用图 6 ～图 8 所述的处理相同。

[0363] 这里，在仅在检测到的发音开始的时刻拍摄图像的情况下，从图 9 所示的语音识别语法中删除作为用于在发音结束时拍摄图像的命令的单词（“Cheese”或“Say Cheese”等）。

[0364] 如果不改变语音识别语法，则改变图 10 所示的识别结果控制数据。将在识别出“Cheese”或“Say Cheese”等时所进行的处理改变成用于在检测到的发音开始的时刻拍摄图像的处理。

[0365] 结果，如果用户说出“Cheese”或“Say Cheese”，则将在发音开始的时刻所拍摄的图像的图像数据存储在存储介质（用于存储图像）111 中。

[0366] 在仅在检测到的发音结束的时刻拍摄图像的情况下，可以类似地进行改变。在这种情况下，省略了下面的处理 :在进行用于检测发音开始的检测操作时拍摄图像的处理（步骤 S605 和 S606）和在取消用于检测发音开始的检测操作时所进行的处理（步骤 S608）。

[0367] 此外，省略由识别结果处理单元 108 所进行的处理中的步骤 S822 ～ S824。

[0368] 这里，如果在步骤 S821 接收到识别结果（步骤 S821 为“是”），则进行步骤 S826 和其后步骤中的处理。

[0369] 此外，从语音识别语法 900 中删除作为用于在发音开始时拍摄图像的命令的单词，或者改变在识别结果控制数据中描述的处理的细节。

[0370] 在第一实施例中，可以将数字照相机 200 配置成：根据识别结果，将在检测到的发音开始的时刻和在检测到的发音结束的时刻所拍摄到的图像的图像数据存储在存储介质（用于存储图像）111 中。

[0371] 例如，如果以在下面的两个时刻拍摄图像的方式描述识别结果控制数据，则将这两个时刻处的图像的图像数据存储在存储介质（用于存储图像）111 中：检测到的“Say Cheese”发音开始的时刻和检测到的“Say Cheese”发音结束的时刻。

[0372] 利用这种结构，可以增加用户期望的摄像时刻的种类数量，并且提高了用户操作的方便性。

[0373] 在第一实施例中，如果在由识别结果处理单元 108 所进行的处理中丢弃识别结果（步骤 S821 为“否”），则用户可以检查是否应该删除存储在存储器（用于存储图像）110 中的图像 A 和 B（步骤 S825）。

[0374] 此外，用户可以选择要被存储在存储介质（用于存储图像）111 中的图像。

[0375] 此外，如果丢弃识别结果，则可以将图像 A 和 B 均存储在存储介质（用于存储图像）111 中。

[0376] 例如，将图像 A 和 B 显示在显示器 115 上，并且可以使用四向选择按钮 204 来选择是否应该删除图像数据。

[0377] 此外，用户使用四向选择按钮 204 选择要存储的图像，并且将在按下确定按钮 205 时所选择的图像的图像数据存储在存储介质（用于存储图像）111 中。

[0378] 如果识别出除作为用于拍摄图像的命令的单词以外的单词（步骤 S826 为“否”），则类似地，用户可以检查是否应该删除图像，并且选择要存储在存储介质（用于存储图像）111 中的图像。

[0379] 此外，可以将图像 A 和 B 的图像数据存储在存储介质（用于存储图像）111 中。

[0380] 利用这种结构，在语音识别性能劣化的环境下应用使用语音命令的摄像功能的情况下，可以防止由于错误识别的语音而错误地删除图像，并且提高了用户操作的方便性。

[0381] 这里，可以根据存储器（用于存储图像）110 的存储容量来确定一个语音识别处理中所保持的图像的数量。

[0382] 利用这种结构，可以考虑存储器（用于存储图像）110 的存储容量，尽可能多地将用户期望的候选图像临时存储在存储器（用于存储图像）110 中。

[0383] 如果在识别结果处理单元 108 所进行的处理中，作为用于在某一时刻拍摄图像的命令的单词的识别得分和作为用于在不同时刻拍摄图像的命令的另一单词的识别得分之间的差小于预定阈值，则可以将在发音开始的时刻和发音结束的时刻所拍摄到的图像都存储在存储介质（用于存储图像）111 中。

[0384] 例如，如果作为用于在发音开始时拍摄图像的命令的“Shoot”的识别得分和作为用于在发音结束时拍摄图像的命令的“Cheese”的识别得分之间的差小于预定值，则将在发

音开始时和发音结束时所拍摄到的图像都存储在存储介质（用于存储图像）111 中。

[0385] 可选地，将这两个图像显示在显示器 115 上，并且用户可以选择其中一个图像或这两个图像。

[0386] 利用这种结构，在语音识别性能可能劣化的环境下应用使用语音命令的摄像功能的情况下，可以防止由于错误识别的语音而错误地删除图像，并且提高了用户操作的方便性。

[0387] 在第一实施例中，对于下面的情况进行了说明：将拍摄到的图像的图像数据临时存储在存储器（用于存储图像）110 中，并且在设置识别结果之后，将图像的图像数据存储在存储介质（用于存储图像）111 中。然而，可以将图像的图像数据直接存储在存储介质（用于存储图像）111 中。

[0388] 在这种情况下，步骤 S608 和 S714 中的用于删除图像数据的处理意为删除存储在存储介质（用于存储图像）111 中的图像数据。

[0389] 此外，不进行步骤 S823 和 S827 中的处理。

[0390] 此外，如果丢弃识别结果（步骤 S821 为“否”），或者如果识别结果不是作为用于拍摄图像的命令的单词（步骤 S826 为“否”），则删除存储在存储介质（用于存储图像）111 中的图像 A 和 B 的图像数据。

[0391] 此外，如果识别结果是作为用于在发音开始时拍摄图像的命令的单词，则删除图像 B 的图像数据。如果识别结果是作为用于在发音结束时拍摄图像的命令的单词，则删除图像 A 的图像数据。

[0392] 例如，在马路边等易受到周围噪声影响的地方使用根据第一实施例的数字照相机 200 的情况下，语音检测单元 106 的内部状态可能在短时间段内频繁改变。

[0393] 如果在短时间段内重复进行图像的拍摄和图像数据的删除，则当启动数字照相机 200 的连续拍摄功能时，数字照相机 200 可能不能在删除图像数据之后立即适当地拍摄图像，并且不能将图像存储在存储器（用于存储图像）110 中。

[0394] 为了解决上述问题，例如，在取消用于检测发音开始的检测操作的时刻，在步骤 S608 不删除拍摄到的图像 A 的图像数据，并且可以将图像 A 的图像数据存储在存储器（用于存储图像）110 中，直到进行用于检测下一发音开始的检测操作的时刻为止。

[0395] 在这种情况下，在进行用于检测下一发音开始的检测操作的时刻，删除图像 A 的图像数据，或者利用新拍摄的图像的图像数据覆盖图像 A 的图像数据。

[0396] 类似地，在步骤 S715 取消用于检测发音结束的检测操作的情况下，可以不删除图像 B 的图像数据，并且可以将其存储在存储器（用于存储图像）110 中，直到进行用于检测下一发音结束的检测操作为止。

[0397] 利用这种结构，即使在进行连续拍摄的速度不快于语音检测状态的改变速度的情况下，也可以至少存储连续拍摄中第一次拍摄的图像。

[0398] 这里，在第一实施例中，对于照相机进行了说明。然而，本发明可应用于摄像机等其它摄像设备。

[0399] 在第一实施例中，使用已知的立体声麦克风作为麦克风 112。

[0400] 此外，语音识别单元 107 可以使用通过左麦克风 112 输入的音频信号的音量和通过右麦克风 112 输入的音频信号的音量之间的关系、或这两个音频信号的音高之间的关系

等，作为上述的特征。

[0401] 通过使用这种特征，例如，可以区分来自数字照相机 200 右侧的声源和来自数字照相机 200 左侧的声源。也就是说，识别拍摄图像时的状况，并且可以拍摄图像。

[0402] 在第一实施例中，代替作为包括在识别结果控制表中的命令的例子所示出的“Cheese”，可以将用于在发音结束时拍摄图像的处理分配给命令“Say Cheese”。

[0403] 此外，代替作为包括在识别结果控制表中的命令的例子所示出的“Go”，可以将用于在发音开始时拍摄图像的处理分配给命令“Now”（好了）。

[0404] 图 16 是示出根据本发明第二实施例的信息处理设备 1600 的结构的例子的功能框图。

[0405] 这里，将以相同的附图标记表示与图 1 所示的组件相同的组件，并且省略对其的说明。

[0406] 可以将信息处理设备 1600 连接到输入设备 1602、摄像设备 1603、存储器设备（用于存储图像）1610、存储设备（用于存储图像）1611 和声音收集器 1612。

[0407] 此外，信息处理设备 1600 可以连接到存储器设备（用于存储语音识别数据）1613、存储器设备（用于存储识别结果控制表）1614 和显示设备 1615。

[0408] 这里，输入设备 1602 具有与操作单元 102 相对应的功能。摄像设备 1603 具有与摄像单元 103 相对应的功能。存储器设备（用于存储图像）1610 具有与存储器（用于存储图像）110 相对应的功能。存储设备（用于存储图像）1611 具有与存储介质（用于存储图像）111 相对应的功能。

[0409] 此外，声音收集器 1612 具有与麦克风 112 相对应的功能。存储器设备（用于存储语音识别数据）1613 具有与存储器（用于存储语音识别数据）113 相对应的功能。

[0410] 此外，存储器设备（用于存储识别结果控制表）1614 具有与存储器（用于存储识别结果控制表）114 相对应的功能。显示控制单元 1609 具有与显示控制单元 109 相对应的功能。

[0411] 信息处理设备 1600 的例子为微处理器等。

[0412] 图 14A 和 14B 以及图 15 是示出由信息处理设备 1600 所进行的处理操作的例子的流程图。

[0413] 首先，使用图 14A 和 14B 的流程图来说明处理。

[0414] 在步骤 S1400，语音输入单元 105 判断是否输入了音频信号。

[0415] 如果没有输入音频信号（步骤 S1400 为“否”），则该过程返回到步骤 S1400。

[0416] 如果输入了音频信号（步骤 S1400 为“是”），则在步骤 S1401，语音检测单元 106 初始化帧 f(f = 0)。

[0417] 接着，在步骤 S1402，语音检测单元 106 将音频信号的检测状态设置为第一状态 301。

[0418] 接着，在步骤 S1403，语音检测单元 106 设置作为检测对象的帧。

[0419] 接着，在步骤 S1404，语音检测单元 106 存储与输入至语音输入单元 105 的音频信号有关的特征数据。

[0420] 这里，特征数据是在语音识别单元 107 进行语音识别时所使用的数据。

[0421] 接着，在步骤 S1405，语音检测单元 106 将语音的检测状态判断为第一状态～第四

状态中的一个。

[0422] 在步骤 S1405，如果语音检测单元 106 将检测状态判断为第一状态 301，则在步骤 S1406，语音检测单元 106 判断作为第一检测是否检测到大于或等于阈值 TH1 的音量。

[0423] 如果检测到大于或等于阈值 TH1 的音量（步骤 S1406 为“是”），则在步骤 S1407，语音检测单元 106 将检测状态改变成第二状态 302（将该时刻称为第一时刻）。

[0424] 接着，在步骤 S1408，摄像控制单元 123 输出用于使得摄像设备 1603 执行摄像操作的信号。

[0425] 这里，根据在步骤 S1408 输出的信号所拍摄到的图像是图像 A。

[0426] 接着，在步骤 S1409，图像存储控制单元 104 输出下面的信号：该信号使得存储器设备（用于存储图像）1610 存储在前一步骤 S1408 中所拍摄到的图像 A 的图像数据，作为第一获取。

[0427] 接着，在步骤 S1410，作为第一存储，语音检测单元 106 存储正被处理的帧 f，作为发音开始帧 Fs。

[0428] 接着，该过程返回至步骤 S1403，并且语音检测单元 106 设置作为下一检测对象的帧。

[0429] 此外，在步骤 S1406，如果没有检测到大于或等于阈值 TH1 的音量（步骤 S1406 为“否”），则该过程同样返回至步骤 S1403，并且语音检测单元 106 设置作为下一语音检测对象的帧。

[0430] 此外，在步骤 S1405，如果语音检测单元 106 将检测状态判断为第二状态 302，则在步骤 S1411，判断正被处理的帧 f 是否是从发音开始帧 Fs 算起的第 M1 个帧或从发音开始帧 Fs 算起的第 M1 个帧之后的帧。

[0431] 此外，如果正被处理的帧 f 在从发音开始帧 Fs 算起的第 M1 个帧之前（步骤 S1411 为“是”），则在步骤 S1413，判断语音检测单元 106 是否检测到小于阈值 TH1 的音量。

[0432] 如果没有检测到小于阈值 TH1 的音量（步骤 S1413 为“否”），则在步骤 S1414，语音检测单元 106 初始化计数器 Fa 的计数值。

[0433] 接着，该过程返回至步骤 S1403，并且语音检测单元 106 设置作为下一语音检测对象的帧。

[0434] 这里，使用计数器 Fa 来判断是否应该复位发音开始帧 Fs。

[0435] 此外，如果检测到小于阈值 TH1 的音量（步骤 S1413 为“是”），则在步骤 S1415，语音检测单元 106 将计数器 Fa 的计数值增大 1。

[0436] 接着，在步骤 S1416，语音检测单元 106 判断计数器 Fa 的计数值是否大于或等于 N1。

[0437] 如果计数器 Fa 的计数值大于或等于 N1（步骤 S1416 为“是”），则在步骤 S1417，图像存储控制单元 104 输出下面的信号：该信号用于删除存储在存储器设备（用于存储图像）1610 中的图像 A 的图像数据。

[0438] 这里，步骤 S1417 中的处理对应于相对用于在进行语音识别之后删除图像数据的处理的第二删除。

[0439] 接着，在步骤 S1418，语音检测单元 106 将检测状态改变成第一状态 301，以再次进行用于检测发音开始的第一检测操作。

[0440] 接着，该过程返回至步骤 S1403，并且语音检测单元 106 设置作为下一语音检测对象的帧。

[0441] 此外，如果计数器 Fa 的计数值小于 N1（步骤 S1416 为“否”），则该过程同样返回至步骤 S1403，并且语音检测单元 106 设置作为下一语音检测对象的帧。

[0442] 此外，在步骤 S1411，如果正被处理的帧 f 是从发音开始帧 Fs 算起的第 M1 个帧或从发音开始帧 Fs 算起的第 M1 个帧之后的帧（步骤 S1411 为“否”），则在步骤 S1412，语音检测单元 106 将检测状态改变成第三状态 303。

[0443] 此外，在步骤 S1405，如果语音检测单元 106 将检测状态判断为第三状态 303，则在步骤 S1419，语音检测单元 106 判断作为第二检测是否检测到小于或等于阈值 TH2 的音量。

[0444] 如果检测到小于或等于阈值 TH2 的音量（步骤 S1419 为“是”），则在步骤 S1420，语音检测单元 106 将检测状态改变成第四状态 304（将该时刻称为第二时刻）。

[0445] 接着，在步骤 S1421，摄像控制单元 123 输出用于使得摄像设备 1603 执行摄像操作的信号。

[0446] 这里，根据步骤 S1421 中输出的信号所拍摄到的图像是图像 B。

[0447] 接着，在步骤 S1422，图像存储控制单元 104 输出下面的信号：该信号用于使得存储器设备（用于存储图像）1610 存储在前一步骤 S1421 所拍摄到的图像 B 的图像数据，作为第二获取。

[0448] 接着，在步骤 S1423，作为第二存储，语音检测单元 106 存储正被处理的帧 f，作为发音结束帧 Fe。

[0449] 接着，该过程返回至步骤 S1403，并且语音检测单元 106 设置作为下一语音检测对象的帧。

[0450] 此外，在步骤 S1419，如果没有检测到小于或等于阈值 TH2 的音量（步骤 S1419 为“否”），则该过程同样返回至步骤 S1403，并且语音检测单元 106 设置作为下一语音检测对象的帧。

[0451] 此外，在步骤 S1405，如果语音检测单元 106 将检测状态判断为第四状态 304，则在步骤 S 1424，判断正被处理的帧 f 是否是从发音结束帧 Fe 算起的第 M2 个帧或从发音结束帧 Fe 算起的第 M2 个帧之后的帧。

[0452] 此外，如果正被处理的帧 f 是从发音结束帧 Fe 算起的第 M2 个帧之前的帧（步骤 S1424 为“是”），则在步骤 S1426，判断语音检测单元 106 是否检测到大于阈值 TH2 的音量。

[0453] 如果没有检测到大于阈值 TH2 的音量（步骤 S1426 为“否”），则在步骤 S1427，语音检测单元 106 初始化计数器 Fb 的计数值。

[0454] 接着，该过程返回至步骤 S1403，并且语音检测单元 106 设置作为下一语音检测对象的帧。

[0455] 这里，使用计数器 Fb 来判断是否应该复位发音结束帧 Fe。

[0456] 此外，如果检测到大于阈值 TH2 的音量（步骤 S1426 为“是”），则在步骤 S1428，语音检测单元 106 将计数器 Fb 的计数值增大 1。

[0457] 接着，在步骤 S1429，语音检测单元 106 判断计数器 Fb 的计数值是否大于或等于 N2。

[0458] 如果计数器 Fb 的计数值大于或等于 N2（步骤 S1429 为“是”），则在步骤 S1430，图

像存储控制单元 104 输出用于删除存储在存储器设备（用于存储图像）1610 中的图像 B 的图像数据的信号。

[0459] 这里，步骤 S1430 中的处理对应于相对用于在进行语音识别之后删除图像数据的处理的第三删除。

[0460] 接着，在步骤 S1431，语音检测单元 106 将检测状态改变成第三状态 303，以再次进行用于检测发音结束的第二检测操作。

[0461] 接着，该过程返回至步骤 S1403，并且语音检测单元 106 设置作为下一语音检测对象的帧。

[0462] 此外，如果计数器 Fb 的计数值小于 N2（步骤 S1429 为“否”），则该过程同样返回至步骤 S1403，并且语音检测单元 106 设置作为下一语音检测对象的帧。

[0463] 此外，如果在步骤 S1424 中正被处理的帧 f 是从发音结束帧 Fe 算起的第 M2 个帧或从发音结束帧 Fe 算起的第 M2 个帧之后的帧（步骤 S1424 为“否”），则在步骤 S1425，语音检测单元 106 结束语音检测。然后该过程进入步骤 S1532。

[0464] 接着，将使用图 15 的流程图来说明处理。

[0465] 在步骤 S1532，语音识别单元 107 根据在步骤 S1504 所获得的帧的特征数据和语音识别数据，进行语音识别。

[0466] 接着，在步骤 S1533，结束由语音识别单元 107 所进行的语音识别。

[0467] 这里，在语音识别单元 107 获得语音识别结果之后，执行步骤 S 1533 中的处理。

[0468] 接着，在步骤 S1534，识别结果处理单元 108 判断识别结果是否表示用于在发音开始时拍摄图像的命令。

[0469] 如果识别结果表示用于在发音开始时拍摄图像的命令（步骤 S1534 为“是”），则在步骤 S1535，输出用于删除图像 B 的信号。

[0470] 如果识别结果不表示用于在发音开始时拍摄图像的命令（步骤 S1534 为“否”），则在步骤 S1536，识别结果处理单元 108 判断语音识别结果是否表示用于在发音结束时拍摄图像的命令。

[0471] 如果识别结果表示用于在发音结束时拍摄图像的命令（步骤 S1536 为“是”），则在步骤 S1537，输出用于删除图像 A 的信号。

[0472] 如果识别结果不表示用于在发音结束时拍摄图像的命令（步骤 S1536 为“否”），则在步骤 S1538，输出用于删除图像 A 和 B 的信号。

[0473] 接着，在步骤 S1539，识别结果处理单元 108 判断识别结果是否表示用于在从发音开始的时刻起过去了特定时间段的时刻拍摄图像的命令。

[0474] 如果识别结果表示用于在从发音开始的时刻起过去了特定时间段的时刻拍摄图像的命令（步骤 S1539 为“是”），则在步骤 S1540，摄像控制单元 123 输出下面的信号：该信号用于使得摄像设备 1603 在过去了特定时间段之后执行摄像操作（将该时刻称为第三时刻）。

[0475] 这里，根据在步骤 S1540 中输出的信号拍摄到的图像是图像 C。

[0476] 接着，在步骤 S1541，图像存储控制单元 104 输出下面的信号：该信号用于使得存储器设备（用于存储图像）1610 存储在前一步骤 S1540 所拍摄到的图像 C 的图像数据，作为第三获取，并且结束该过程。

[0477] 此外，如果识别结果不表示用于在从发音开始的时刻起过去了特定时间段的时刻拍摄图像的命令（步骤 S1539 为“否”），则结束该过程。

[0478] 利用这种结构，在发音期间，可以获得作为第一关系的在发音开始时所拍摄的第一图像（图像 A）和作为第二关系的在发音结束时所拍摄的第二图像（图像 B）。

[0479] 此外，在发音期间可以获得作为第三关系的在从发音开始起过去了特定时间段的时刻所拍摄到的第三图像（图像 C）。

[0480] 此外，根据发音期间内的语音内容，可以从多个图像中选择在用户期望的时刻所拍摄到的图像。

[0481] 此外，利用这种结构，通过与根据第二实施例的信息处理设备 1600 同步地操作外部装置，可以高效获得在用户期望的时刻所拍摄到的图像。

[0482] 此外，根据按照第二实施例的信息处理设备 1600，即使在输入断续语音的情况下，也可以将这种断续语音识别为一个命令。因此，即使在使用发音期间长的单词作为命令的情况下，也降低了识别错误的可能性。

[0483] 这里，还可以通过向系统或设备提供存储有实现上述实施例所述功能的软件的程序代码的存储介质，并且通过由该系统或设备的计算机读取并执行该程序代码，来实现本发明。

[0484] 这里，计算机可以是中央处理单元 (CPU) 或微处理器单元 (MPU) 等。

[0485] 在这种情况下，作为计算机可读的且从存储介质读取的程序代码实现上述实施例所述的功能。存储该程序代码的存储介质为本发明。

[0486] 用于提供程序代码的存储介质的例子有软盘、硬盘、光盘、磁光盘、紧凑型光盘只读存储器 (CD-ROM)、可记录紧凑型光盘 (CD-R)、磁带、非易失性存储卡和只读存储器 (ROM) 等。

[0487] 此外，不是必须仅通过执行由计算机所读取的程序代码才能实现上述实施例所述的功能。操作系统 (OS) 等可以根据程序代码的内容进行用于实现上述实施例所述功能的部分或全部实际处理。

[0488] 这里，本发明还包括通过该处理实现上述实施例所述的功能的情况。

[0489] 这里，OS 运行在计算机上。

[0490] 此外，将从存储介质读取的程序代码写入包括在插入计算机的功能扩展板内的存储器中或写入包括在与计算机连接的功能扩展单元内的存储器中。

[0491] 本发明还包括下面的情况：此后，包括在功能扩展板或功能扩展单元中的 CPU 根据程序代码的内容，进行部分或全部实际处理，并且通过该处理实现上述实施例所述的功能。

[0492] 尽管已经参考典型实施例说明了本发明，但是应该理解，本发明不局限于所公开的典型实施例。所附权利要求书的范围符合最宽的解释，以包含所有这类修改、等同结构和功能。

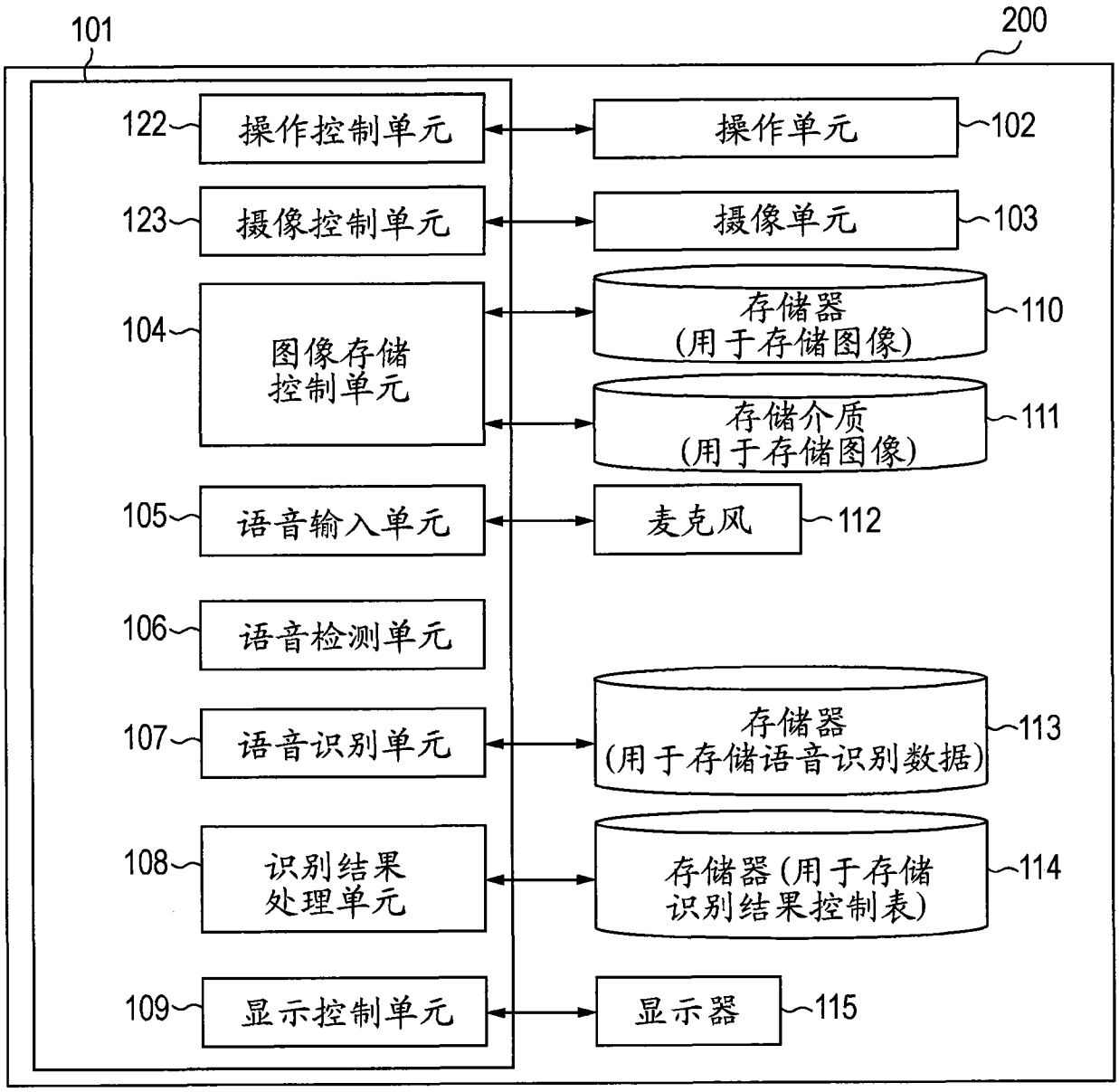

101
200
122
操作控制单元
操作单元
102
123
摄像控制单元
摄像单元
103
104
图像存储
控制单元
存储器
(用于存储图像)
110
存储介质
(用于存储图像)
111
105
语音输入单元
麦克风
112
106
语音检测单元
107
语音识别单元
存储器
(用于存储语音识别数据)
113
108
识别结果
处理单元
存储器(用于存储
识别结果控制表)
114
109
显示控制单元
显示器
115

图 1

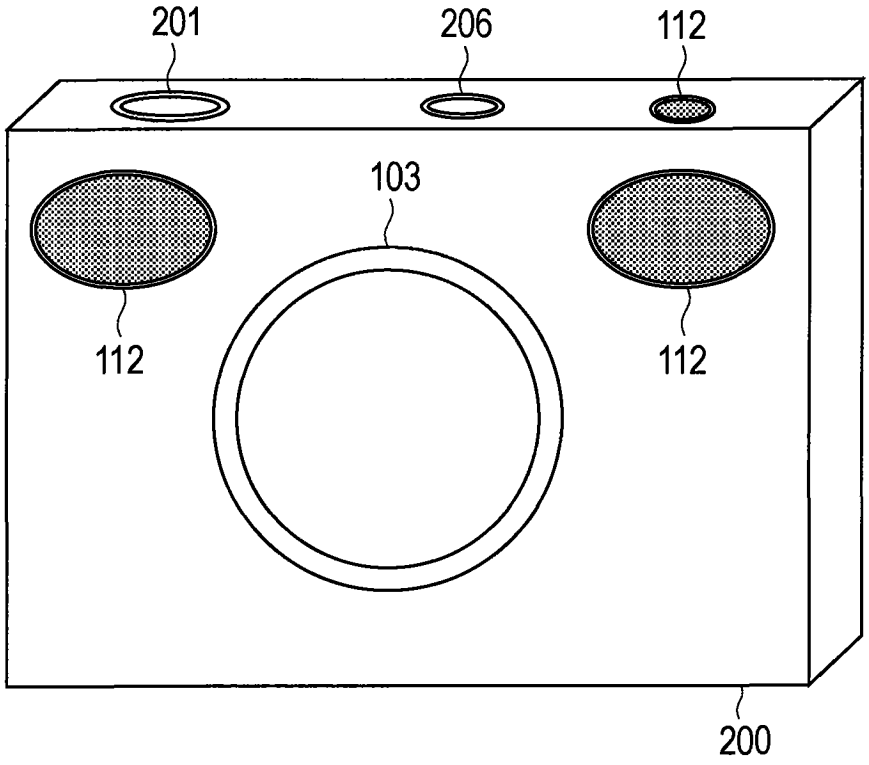

201
206
112
103
112
112
200

图 2A

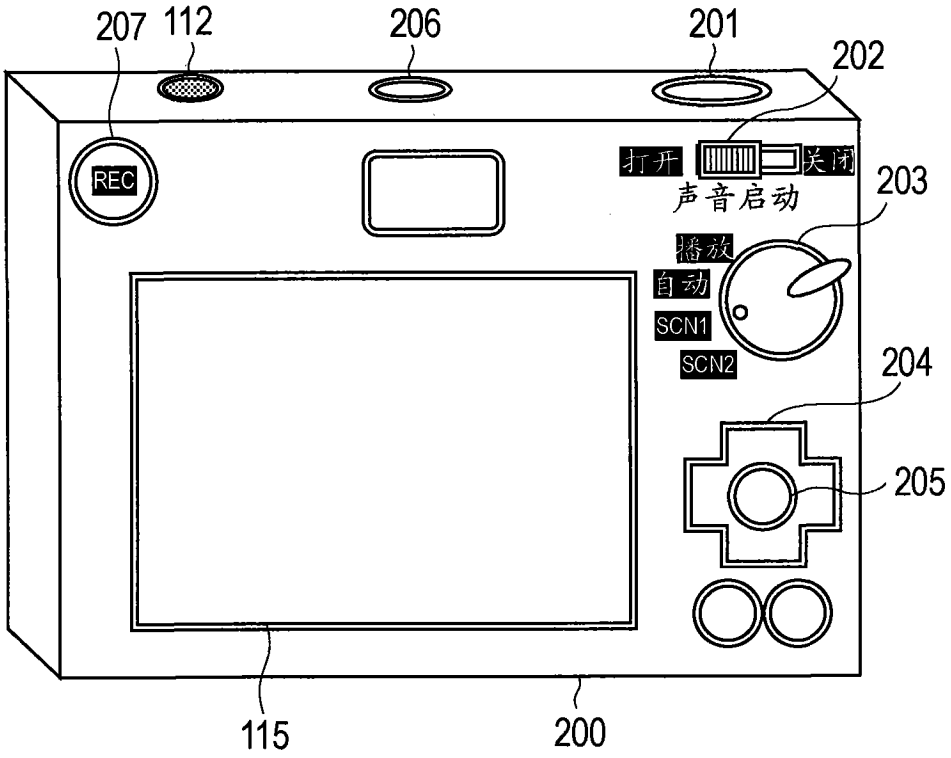

207
112
206
201
202
REC
打开
关闭
声音启动
203
播放
自动
SCN1
SCN2
204
205
115
200

图 2B

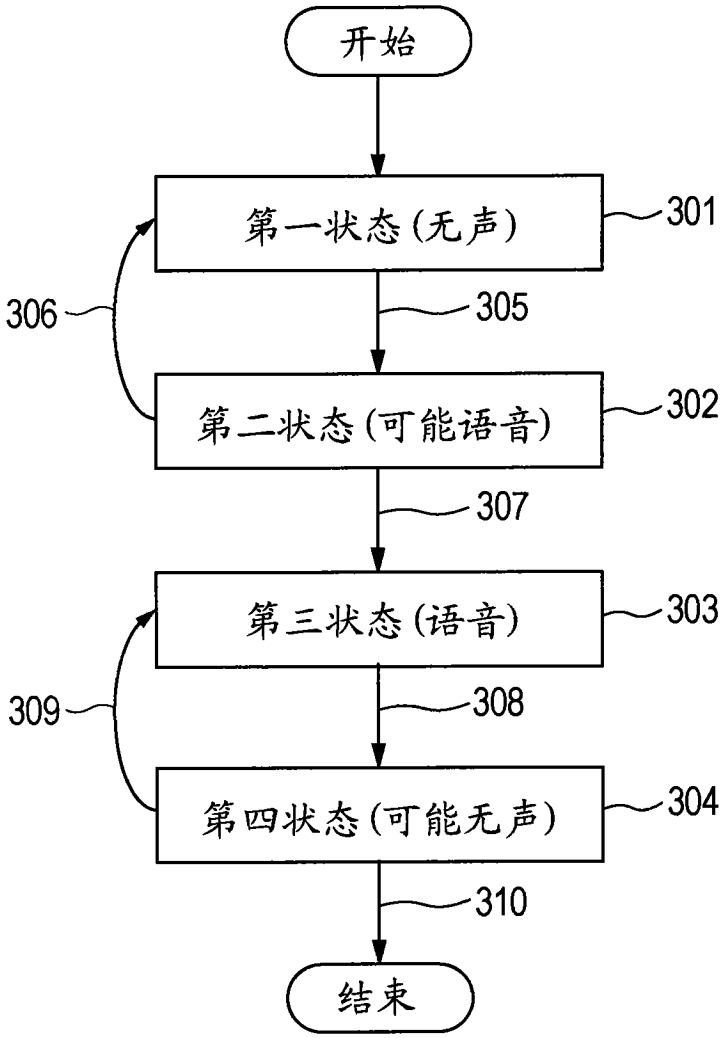

开始
第一状态(无声)
301
306
305
第二状态(可能语音)
302
307
第三状态(语音)
303
309
308
第四状态(可能无声)
304
310
结束

图 3

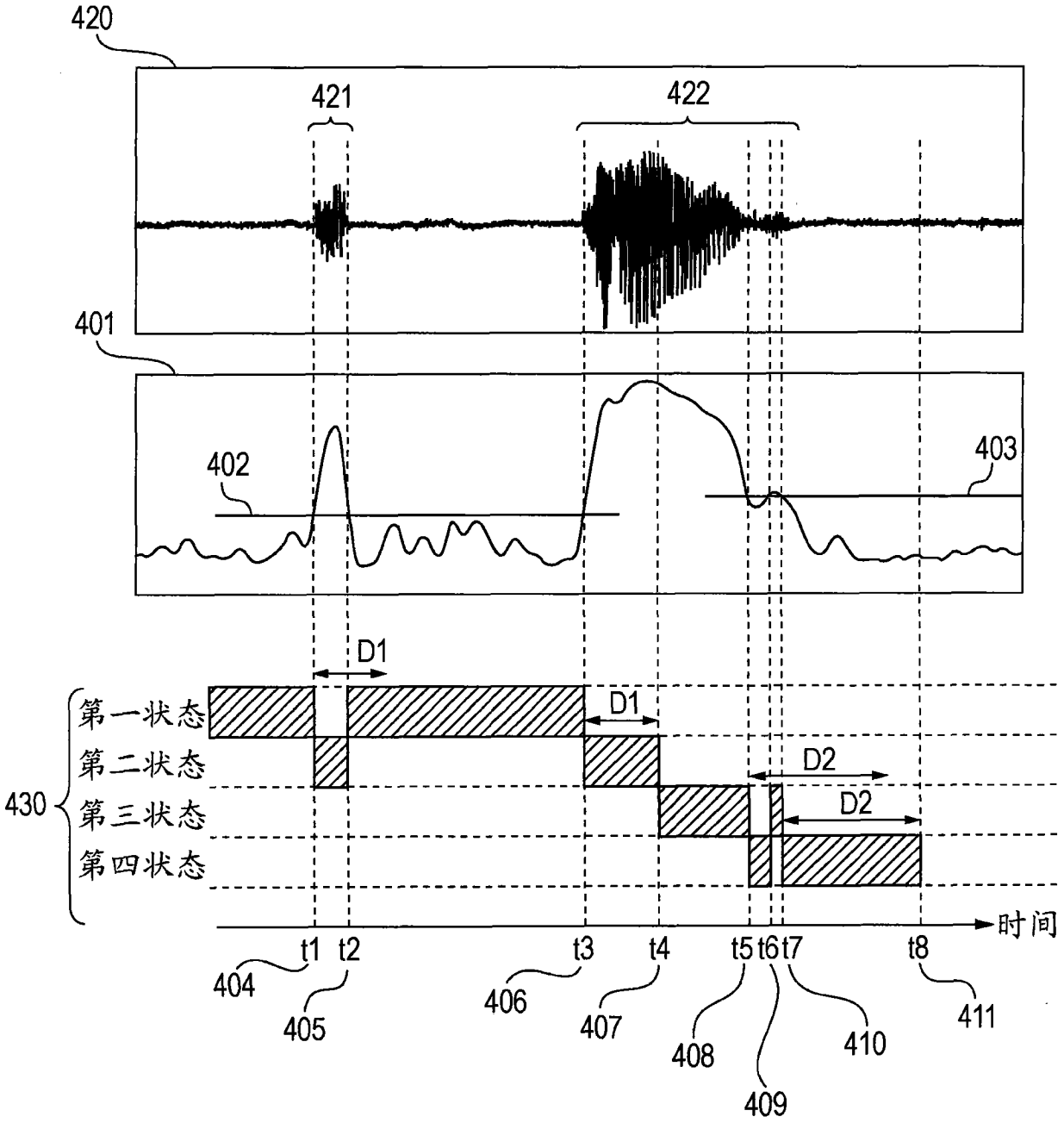

420
421
422
401
402
403
D1
D1
D2
D2
第一状态
第二状态
第三状态
第四状态
430
时间
t1
t2
t3
t4
t5
t6
t7
t8
404
405
406
407
408
409
410
411

图 4

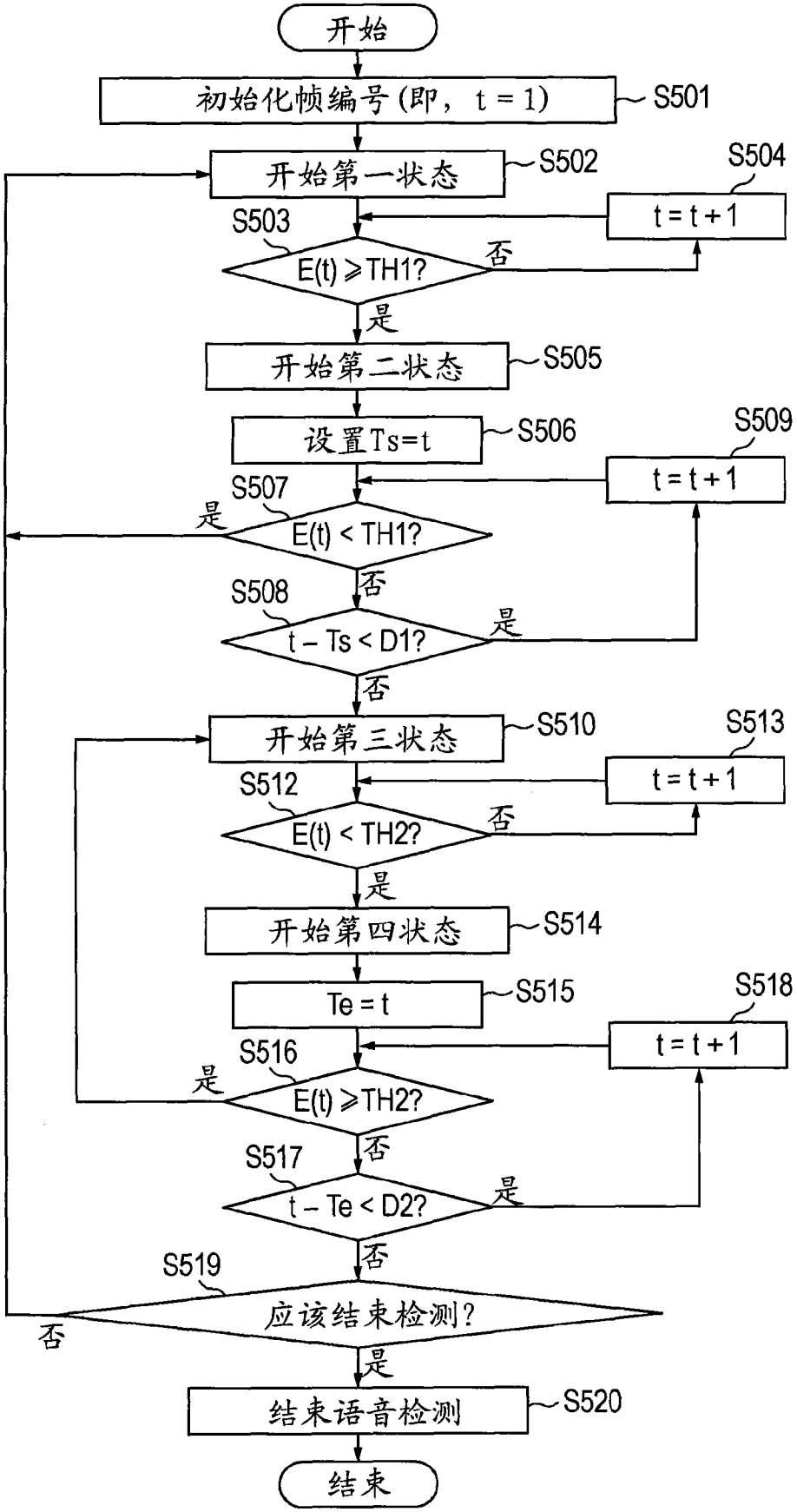

开始
初始化帧编号(即，t = 1)
S501
开始第一状态
S502
S503
E(t) ≥ TH1?
否
是
S504
t = t + 1
开始第二状态
S505
设置Ts=t
S506
S509
t = t + 1
S507
是
E(t) < TH1?
否
S508
t – Ts < D1?
是
否
开始第三状态
S510
S513
t = t + 1
S512
E(t) < TH2?
否
是
开始第四状态
S514
Te = t
S515
S518
t = t + 1
S516
是
E(t) ≥ TH2?
否
S517
t – Te < D2?
是
否
S519
应该结束检测?
否
是
结束语音检测
S520
结束

图 5

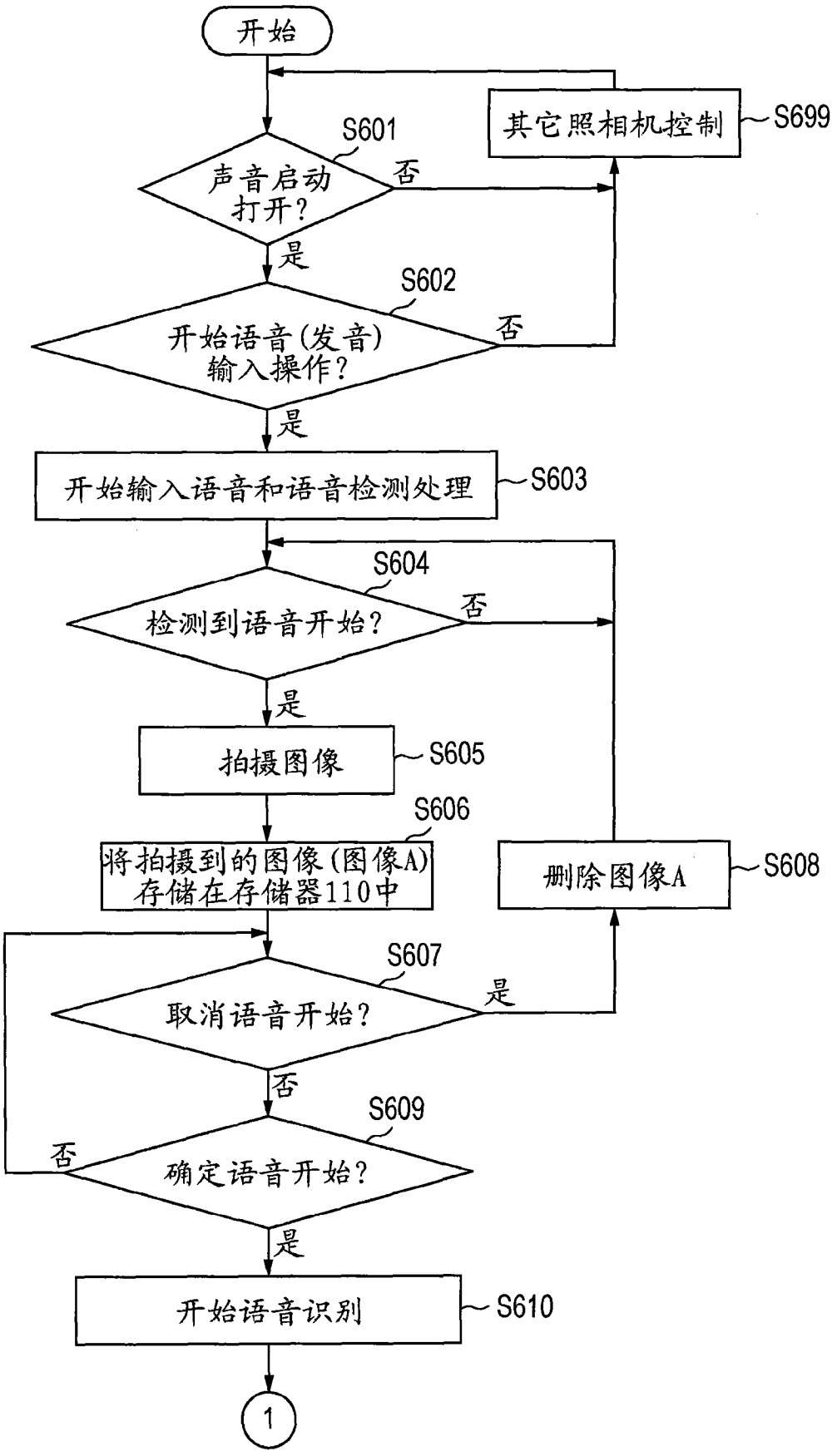

开始
S601
声音启动打开?
否
是
其它照相机控制
S699
S602
开始语音(发音)输入操作?
否
是
开始输入语音和语音检测处理
S603
S604
检测到语音开始?
否
是
拍摄图像
S605
S606
将拍摄到的图像(图像A)存储在存储器110中
删除图像A
S608
S607
取消语音开始?
是
否
S609
确定语音开始?
否
是
开始语音识别
S610
1

图 6

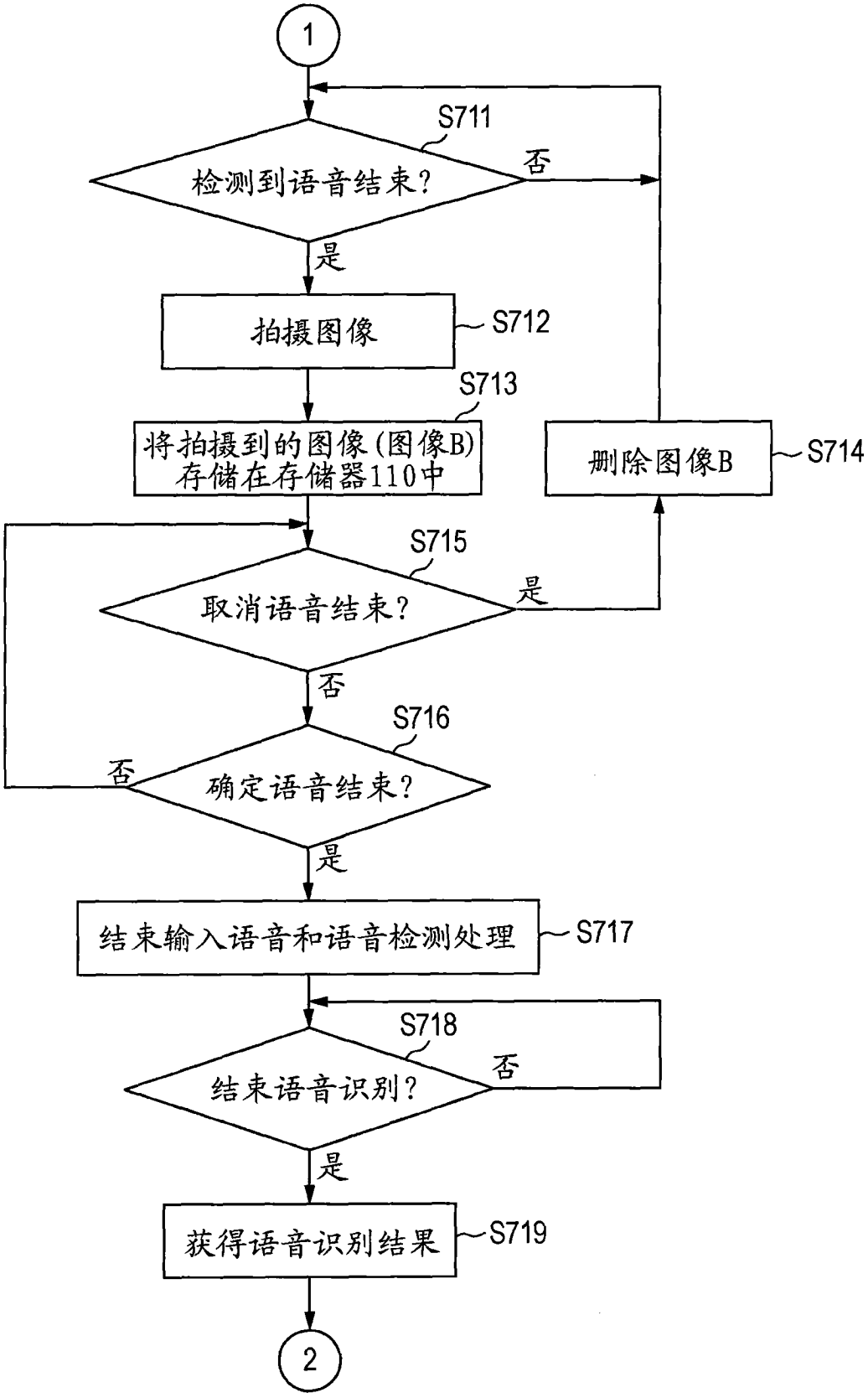

1
S711
检测到语音结束?
否
是
拍摄图像
S712
S713
将拍摄到的图像(图像B)存储在存储器110中
删除图像B
S714
S715
取消语音结束?
是
否
S716
确定语音结束?
否
是
结束输入语音和语音检测处理
S717
S718
结束语音识别?
否
是
获得语音识别结果
S719
2

图 7

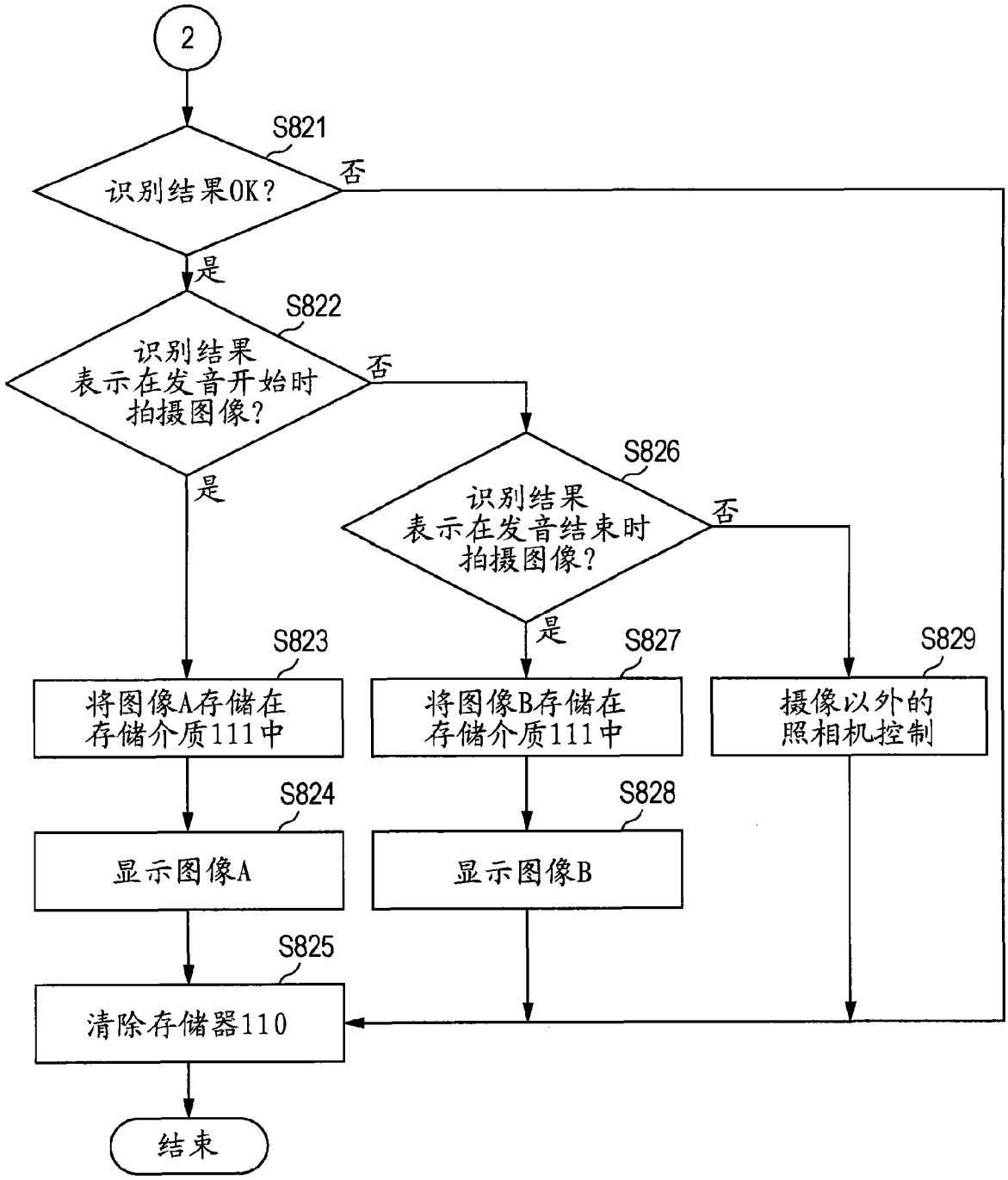

2
S821
识别结果OK?
否
是
S822
识别结果
表示在发音开始时
拍摄图像?
否
是
S826
识别结果
表示在发音结束时
拍摄图像?
否
是
S823
将图像A存储在
存储介质111中
S827
将图像B存储在
存储介质111中
S829
摄像以外的
照相机控制
S824
显示图像A
S828
显示图像B
S825
清除存储器110
结束

图 8

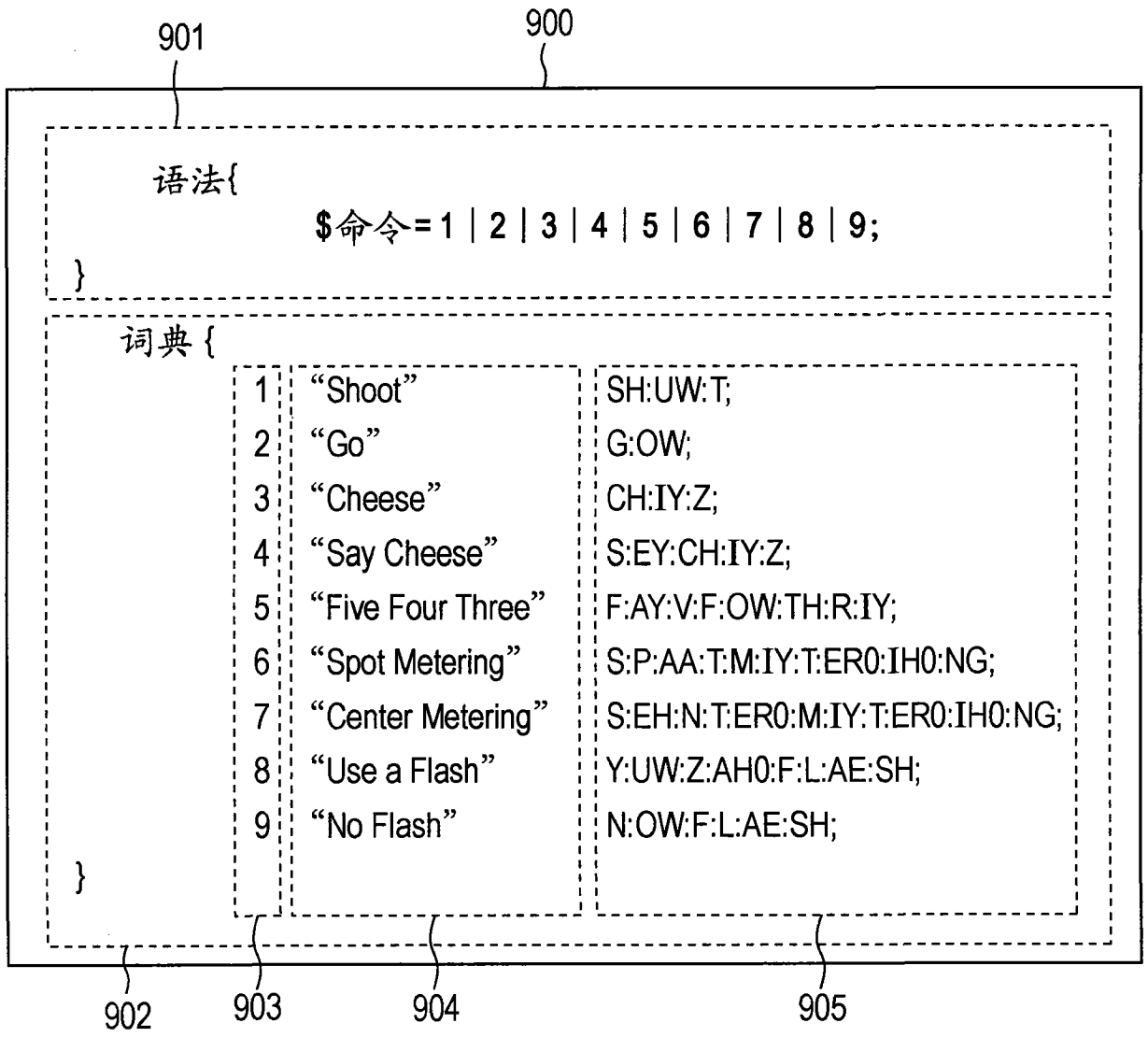

901
900
语法{
$命令=1|2|3|4|5|6|7|8|9;
}
词典{
1 "Shoot" SH:UW:T;
2 "Go" G:OW;
3 "Cheese" CH:IY:Z;
4 "Say Cheese" S:EY:CH:IY:Z;
5 "Five Four Three" F:AY:V:F:OW:TH:R:IY;
6 "Spot Metering" S:P:AA:T:M:IY:T:ER0:IH0:NG;
7 "Center Metering" S:EH:N:T:ER0:M:IY:T:ER0:IH0:NG;
8 "Use a Flash" Y:UW:Z:AH0:F:L:AE:SH;
9 "No Flash" N:OW:F:L:AE:SH;
}
902
903
904
905

图 9

命令	处理
Shoot	保持在发音开始时所拍摄到的图像
Go	保持在发音开始时所拍摄到的图像
Cheese	保持在发音结束时所拍摄到的图像
Say Cheese	保持在从发音结束的时刻起过去了特定时间段的时刻所拍摄到的图像
Five Four Three	保持在从发音开始的时刻起过去了特定时间段的时刻所拍摄到的图像
Spot Metering	启动点测光
Center Metering	启动中央重点测光
Use a Flash	启动闪光灯
No Fash	禁用闪光灯

904 1002

图 10

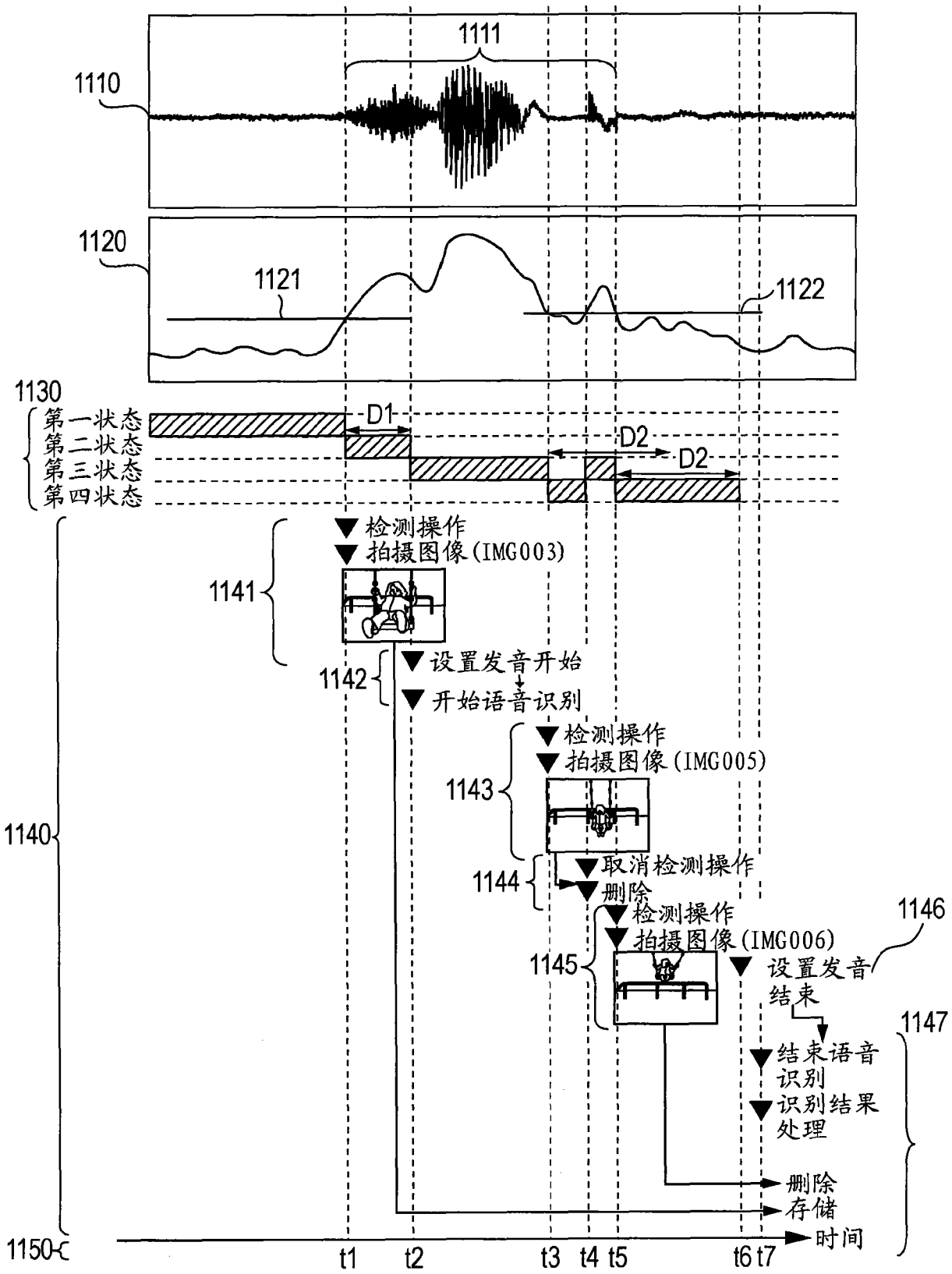

1110
1111
1120
1121
1122
1130
第一状态
第二状态
第三状态
第四状态
D1
D2
D2
1140
1141
检测操作
拍摄图像(IMG003)
1142
设置发音开始
开始语音识别
1143
检测操作
拍摄图像(IMG005)
1144
取消检测操作
删除
1145
检测操作
拍摄图像(IMG006)
1146
设置发音结束
1147
结束语音识别
识别结果处理
删除
存储
1150
时间
t1
t2
t3
t4
t5
t6
t7

图 11

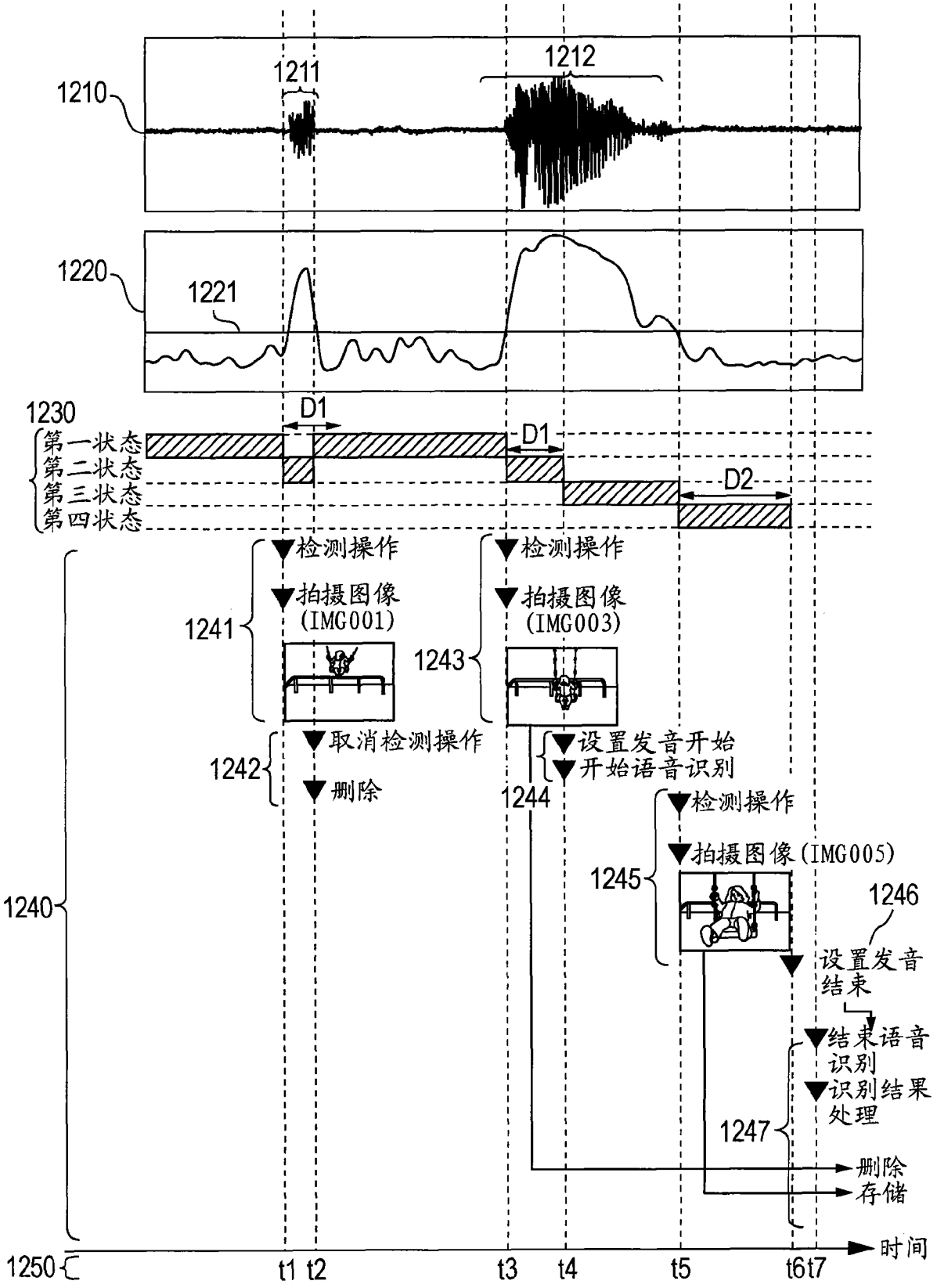

1210
1211
1212
1220
1221
1230
第一状态
第二状态
第三状态
第四状态
D1
D1
D2
1240
1241
检测操作
拍摄图像
(IMG001)
1242
取消检测操作
删除
1243
检测操作
拍摄图像
(IMG003)
1244
设置发音开始
开始语音识别
1245
检测操作
拍摄图像(IMG005)
1246
设置发音结束
1247
结束语音识别
识别结果处理
删除
存储
1250
t1 t2 t3 t4 t5 t6t7
时间

图 12

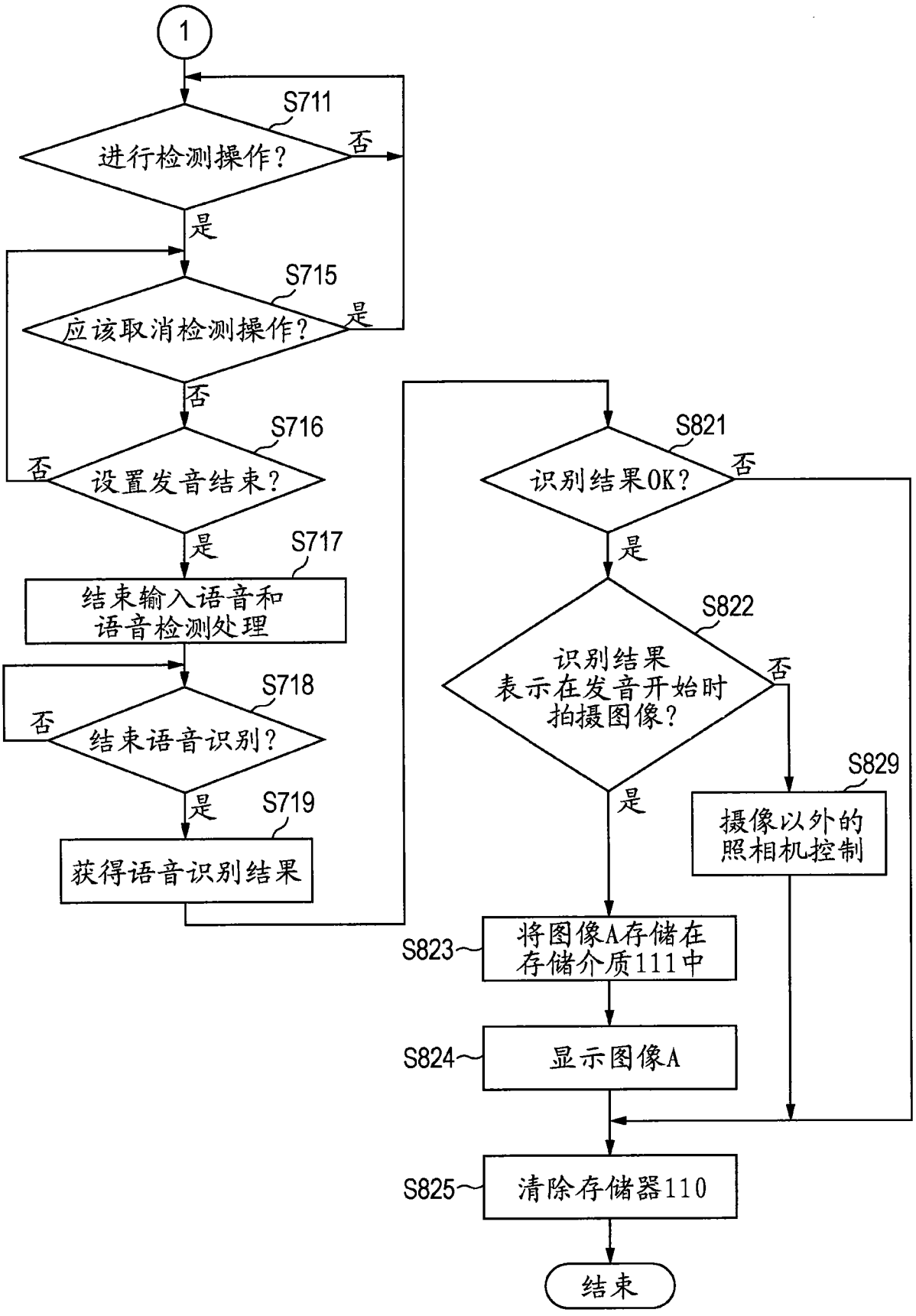

1
S711
进行检测操作?
否
是
S715
应该取消检测操作?
是
否
S716
设置发音结束?
否
是
S717
结束输入语音和
语音检测处理
S718
结束语音识别?
否
是
S719
获得语音识别结果
S821
识别结果OK?
否
是
S822
识别结果
表示在发音开始时
拍摄图像?
否
是
S829
摄像以外的
照相机控制
S823
将图像A存储在
存储介质111中
S824
显示图像A
S825
清除存储器110
结束

图 13

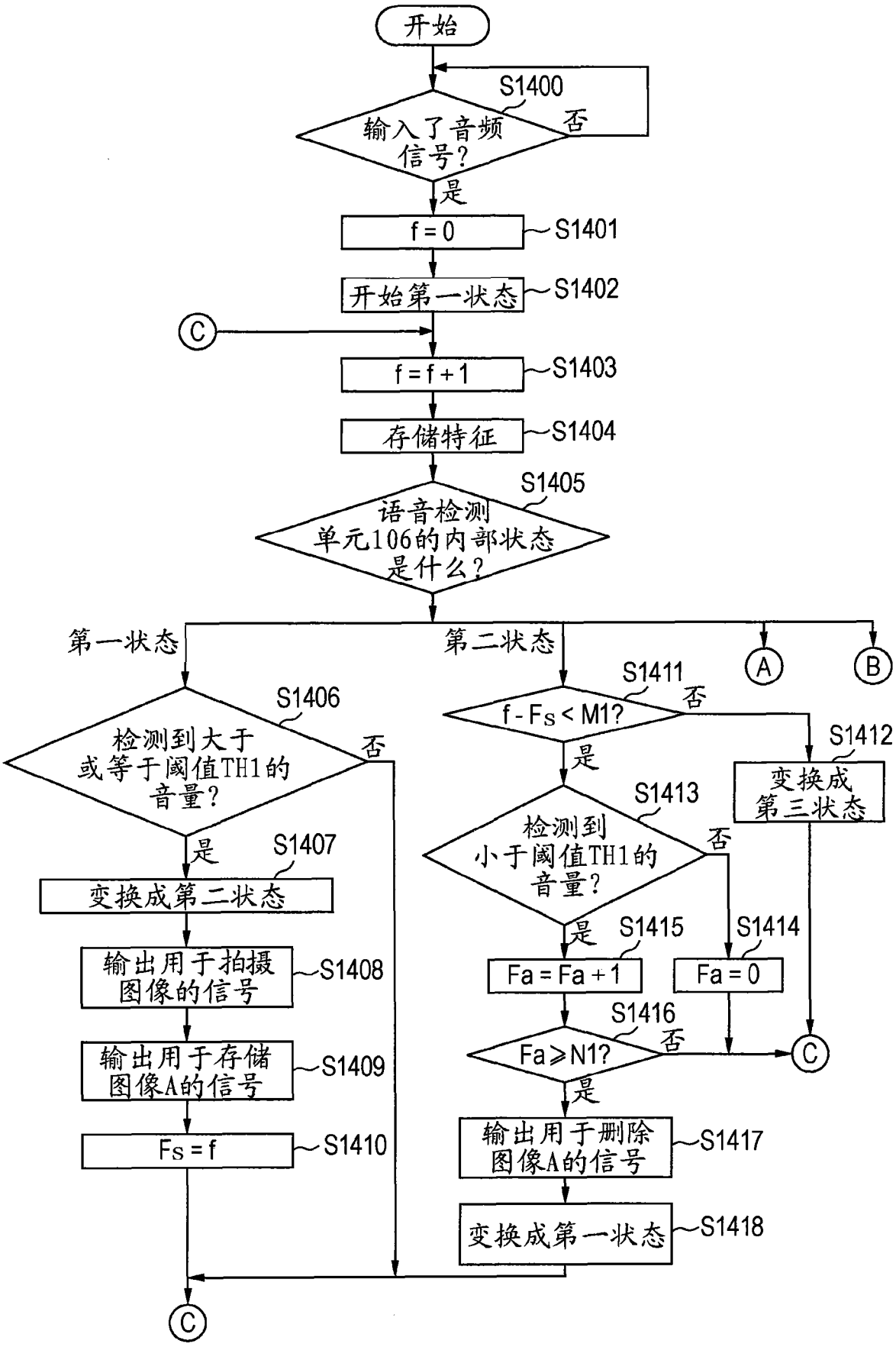

开始
S1400
输入了音频信号?
否
是
f = 0
S1401
开始第一状态
S1402
C
f = f + 1
S1403
存储特征
S1404
S1405
语音检测单元106的内部状态是什么?
第一状态
第二状态
A
B
S1406
检测到大于或等于阈值TH1的音量?
否
是
S1407
变换成第二状态
输出用于拍摄图像的信号
S1408
输出用于存储图像A的信号
S1409
Fs = f
S1410
C
S1411
f - Fs < M1?
否
是
S1412
变换成第三状态
S1413
检测到小于阈值TH1的音量?
否
是
S1415
Fa = Fa + 1
S1414
Fa = 0
S1416
Fa ≥ N1?
否
是
C
输出用于删除图像A的信号
S1417
变换成第一状态
S1418

图 14A

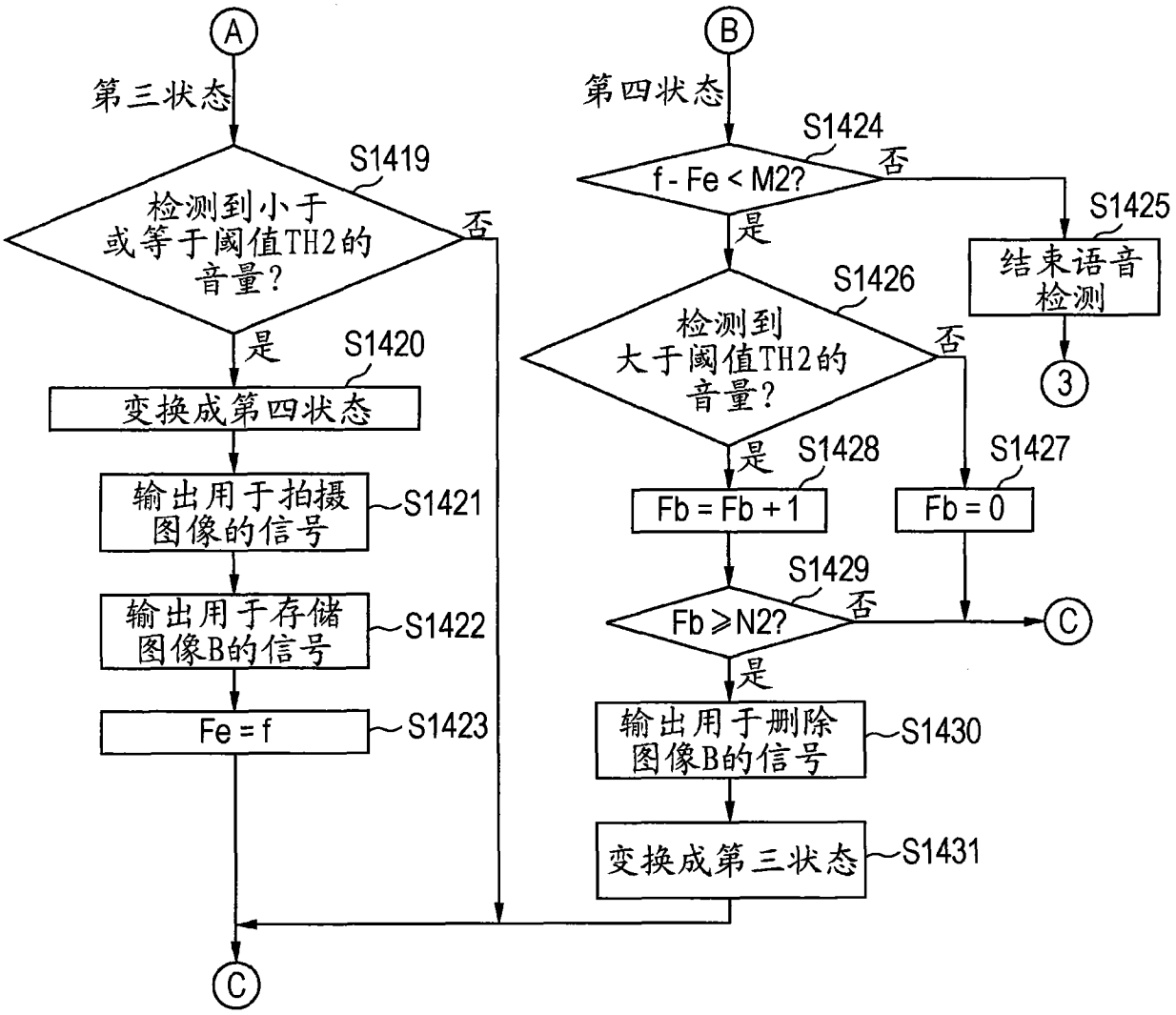

A
第三状态
S1419
检测到小于或等于阈值TH2的音量?
否
是
S1420
变换成第四状态
输出用于拍摄图像的信号
S1421
输出用于存储图像B的信号
S1422
Fe = f
S1423
C
B
第四状态
S1424
f - Fe < M2?
否
是
S1425
结束语音检测
3
S1426
检测到大于阈值TH2的音量?
否
是
S1428
Fb = Fb + 1
S1427
Fb = 0
S1429
Fb ≥ N2?
否
C
是
输出用于删除图像B的信号
S1430
变换成第三状态
S1431

图 14B

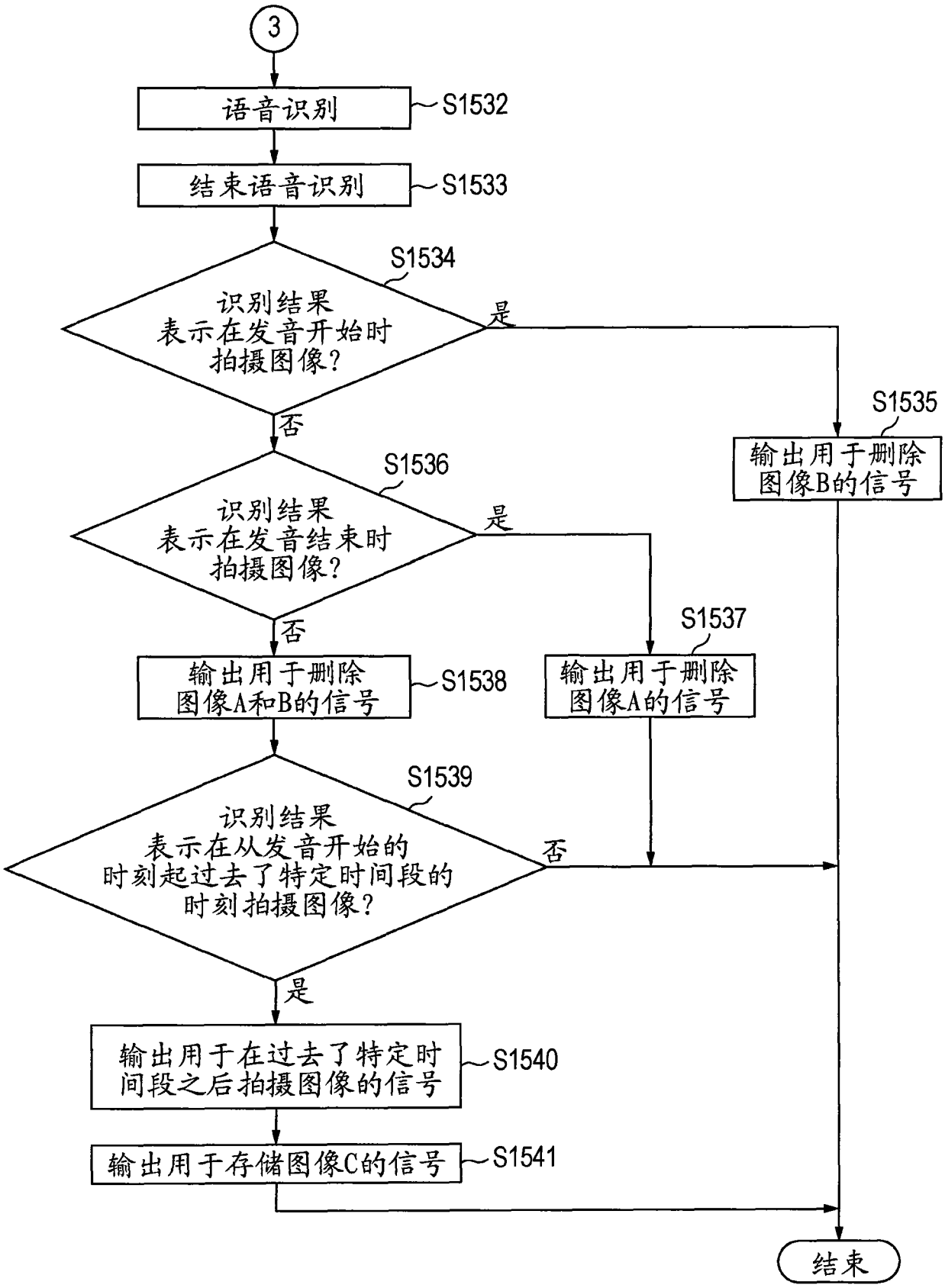

3
语音识别
S1532
结束语音识别
S1533
S1534
识别结果
表示在发音开始时
拍摄图像?
是
否
S1535
输出用于删除
图像B的信号
S1536
识别结果
表示在发音结束时
拍摄图像?
是
否
S1537
输出用于删除
图像A的信号
输出用于删除
图像A和B的信号
S1538
S1539
识别结果
表示在从发音开始的
时刻起过去了特定时间段的
时刻拍摄图像?
否
是
输出用于在过去了特定时
间段之后拍摄图像的信号
S1540
输出用于存储图像C的信号
S1541
结束

图 15

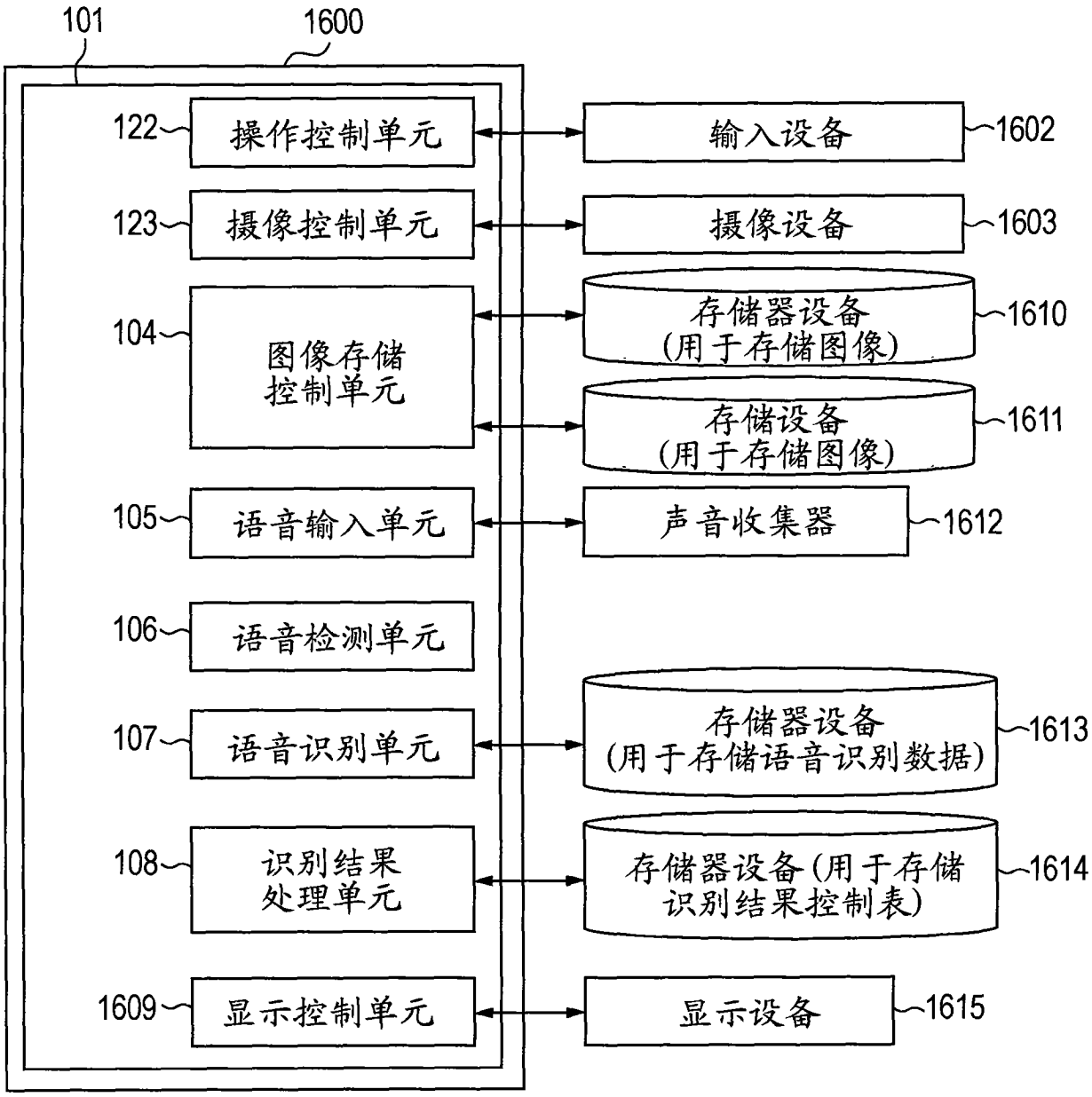

101
1600
122
操作控制单元
123
摄像控制单元
104
图像存储
控制单元
105
语音输入单元
106
语音检测单元
107
语音识别单元
108
识别结果
处理单元
1609
显示控制单元
输入设备
1602
摄像设备
1603
存储器设备
(用于存储图像)
1610
存储设备
(用于存储图像)
1611
声音收集器
1612
存储器设备
(用于存储语音识别数据)
1613
存储器设备(用于存储
识别结果控制表)
1614
显示设备
1615

图 16