



[12] 发明专利申请公开说明书

[21] 申请号 02127545.9

[43] 公开日 2003 年 1 月 15 日

[11] 公开号 CN 1391210A

[22] 申请日 2002.5.8 [21] 申请号 02127545.9

[30] 优先权

[32] 2001.5.11 [33] DE [31] 10122828.7

[71] 申请人 皇家飞利浦电子有限公司

地址 荷兰艾恩德霍芬

[72] 发明人 V·施泰恩比斯 T·埃塞勒

[74] 专利代理机构 中国专利代理(香港)有限公司

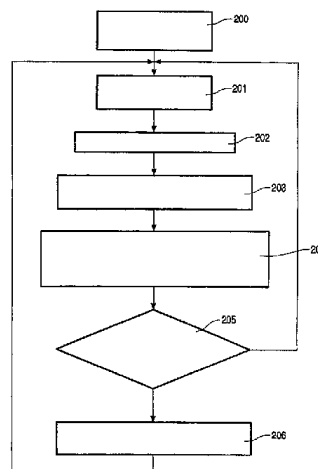
代理人 张志醒

权利要求书 3 页 说明书 8 页 附图 5 页

[54] 发明名称 用于一种语言识别设备的训练或适配方法

[57] 摘要

本发明涉及用于遵照一个电子设备的功能行事的一个语音识别设备的训练或适配,例如在一个移动电话中一个语音拨号的触发。为了构造语音识别设备训练和/或适配以便改善用户舒适性,提出具有下面步骤的一种方法:执行一个语音输入;处理由语音识别设备输入的语音以便产生一个语音识别结果;如果该语音识别结果能够被分配到电子设备的一个功能,遵照该电子设备的所分配的功能行事;如果遵照电子设备的所分配的功能行事不能导致一个表示拒绝的用户输入,以与完成的语音输入关联的语音识别结果为基础训练或适配该语音识别设备。



1. 一种用于遵照一个电子设备(1)的功能行事的一个语音识别设备(4)的训练或适配的方法, 其中提供下列步骤:

5 执行一个语音输入;

 处理由语音识别设备(4)输入的语音以便产生一个语音识别结果;

 如果该语音识别结果能够被分配给电子设备(1)的一个功能, 遵照该电子设备(1)的所分配的功能行事;

10 如果遵照电子设备(1)的所分配的功能行事不能导致一个表示拒绝的用户输入, 以与完成的语音输入关联的语音识别结果为基础训练或适配该语音识别设备(4)。

2. 如权利要求1所述的方法, 其特征在于, 如果电子设备(1)的功能的运行没有被一个用户输入中断, 语音识别设备(4)的训练和适配发生。

15 3. 如权利要求1或2所述的方法, 其特征在于, 对于该语音识别设备(4)的一个语音识别结果, 一个可靠性级别被确定, 并且如果该语音识别结果的可靠性级别落在一个预定的目标范围内, 以对应于该完成的语音输入的语音识别结果为基础执行该语音识别设备(4)的一个训练或适配。

4. 如权利要求1或2所述的方法, 其特征在于,

20 对于一个语音识别设备(4)的语音识别结果来说, 确定一个最初的可靠性级别;

 如果该语音识别结果的最初的可靠性级别落在一个预定的目标范围内, 以对应于该完成的语音输入的语音识别结果为基础执行该语音识别设备的(4)的训练或适配,

25 如果这个初始的可靠性级别没有落在一个最初指定的目标范围内, 并且同样如果通过另一种输入方法完成随后的用户输入, 则考虑到通过其它输入方法完成的输入, 它提供一个对于该语音识别结果被确定的第二可靠性级别, 并且如果这个第二可靠性级别落在一个第二指定的目标范围内, 以考虑到通过另一种输入方法完成的用户输入的语音识别结果为基础执行训练操作或适配。

5. 如权利要求 1 或 2 所述的方法，其特征在于，如果在一个语音输入之后进行一个表示拒绝该语音识别结果的用户输入，并且随后地通过另一种输入方法完成一个用户输入，则考虑到通过其它输入方法完成的输入，它提供对于该语音识别结果被确定的第二可靠性级别，并且执行语音识别设备(4)的训练操作或适配，提供语音输入和表示该相同的词语或词语序列的通过其它输入方法完成的用户输入。

6. 如权利要求 1 或 2 所述的方法，其特征在于，

该语音识别设备（4）在每种情况下也提供至少一个可替换的语音识别结果，

10 如果遵照分配给语音识别结果的电子设备的所功能行事不能导致一个表示该语音识别结果的拒绝的用户输入，以与最后的语音输入关联的语音识别结果为基础执行该语音识别设备（4）的训练或适配，并且如果

该语音识别结果对于最后之前的语音输入对应于一个可替换的语音识别结果并且

15 在最后之前进行一个表示语音识别结果的拒绝的用户输入，

执行语音识别设备的一个进一步的训练操作或进一步的适配，提供最后的语音识别结果和表示该相同的词语或词语序列的最后之前的一个。

7. 如权利要求 1 至 6 的其中之一要求的方法，其特征在于提供一个语音识别结果的听觉上或视觉上的再现。

20 8. 如权利要求 1 至 7 之一所述的方法，其特征在于，该电子设备（1）是一个移动电话。

9. 如权利要求 8 所述的方法，其特征在于，通过一个语音输入触发一个拨号过程。

25 10. 一种具有语音识别设备（4）的电子设备（1），用于遵照该电子设备的功能行事，其中该电子设备（1）和语音识别设备（4）以这样一种方式设计，下面的步骤被用于该语音识别设备（4）的训练或适配：

执行一个语音输入；

处理由语音识别设备（4）输入的语音以便产生一个语音识别结果；

如果该语音识别结果能够被分配到电子设备（1）的一个功能，遵照

该电子设备（1）的所分配的功能行事；

如果遵照电子设备（1）的所分配的功能行事不能导致一个表示拒绝的用户输入，以与完成的语音输入关联的语音识别结果为基础训练或适配该语音识别设备（4）。

- 5 11. 一种具有语音识别设备（4）的移动电话终端（1），用于遵照该移动电话终端的功能行事，其中该移动电话终端（1）和语音识别设备（4）以这样一种方式设计，下面的步骤被用于该语音识别设备（4）的训练或适配：

 执行一个语音输入；

 处理由语音识别设备（4）输入的语音以便产生一个语音识别结果；

- 10 如果该语音识别结果能够被分配到移动电话终端（1）的一个功能，遵照该移动电话终端（1）的所分配的功能行事；

 如果遵照移动电话终端（1）的所分配的功能行事不能导致一个表示拒绝的用户输入，以与完成的语音输入关联的语音识别结果为基础训练或适配该语音识别设备（4）。

理为基础应用到那里；然而该参数也能够与用户无关地被预置。在于该用户的
声音适配中，一个系统通过收集这个用户的特殊语音数据被改进，也就是在适
配期间的误差率将降低。许多种适配方法是已知的，例如MAP（最大后验法），
MLLR（最大似然线性回归），专用声音（这种方法目前只用在复杂的结构系
5 统中）。同样除声音适配之外的适配形式是可能的；因此系统性能和这些数据
的分析的改进能被实现，例如通过建立使用率统计（例如通过估计多久一条指
令或电话号码被讲一次，并且结合它们所讲的内容）。

US5,805,672 公开了通过具有一个语音识别设备的辅助设备实现在机动车
辆中移动电话设备的语音拨号功能的设备。当在机动车辆中使用时，一个免提
10 装置被使用。如果需要语音拨号，则语音拨号模式必须首先被激活。该辅助
设备然后通过语音识别设备识别在语音输入中已学过的单词，并给一个已识别
的单词指定一个相关的电话号码，然后用户相关的发音被自动地拨叫。一种训练
模式被提供，通过它的激活，被训练的语音识别设备能够被训练。这个训练模
式由压下按键而被激活。在训练期间，用户被要求说出一个名字两次并输入相
15 关的电话号码。接着用户按下一个确定键，在此之后该系统通过语音输出向该
用户输出该语音识别结果（识别的名字）和相关的电话号码。该用户能够继续
另一个名字的训练。

WO00/14729 公开了一种具有语音识别设备和语音拨号功能的移动电话。
在这里也提供一种由开关操作起动的训练模式。在该训练模式期间，一个
20 普通的训练操作被执行，其中通过新单词的输入扩展语音识别设备的词汇。

发明内容

本发明是基于设计该语音识别设备的训练和 / 或适配以便增加用户舒适性
这个目的的。

这个目的用最初描述的方法通过下列步骤实现：

- 25
- 一执行语音输入；
 - 一通过语音识别设备处理该语音输入以便产生一个语音识别结果；
 - 一如果该语音识别能被指配为该电子设备的一个功能，则完成该电子设
备的指定功能；
 - 一如果对电子设备的指配功能的动作没有导致一个用户引起拒绝，则以

用于一种语音识别设备的训练或适配方法

5 技术领域

本发明涉及一种用于一种用于遵照一个电子设备的功能行事的语音识别设备的训练或适配方法。

背景技术

特别地，该电器设备是一个移动电话，其中语音拨号的性能与所实现的设备功能相一致，其中通过一个语音输入另一个用户被拨叫。当用在一个机动车辆中时在其中通常提供免提装置以进一步辅助用户，因为他不再执行键盘输入以触发对另一个用户的拨叫，用户的操作舒适性被改进。应用的其它领域包括诸如电视接收机，录像机，高保真音乐系统厨房设备或具有语音控制的其它设备的电子用户设备的控制，其中各种设备功能能够通过语音执行，例如音量控制，电视频道或无线电台的选择，或厨房设备的发动机转速的调整。

一种语音识别设备能够被训练和 / 或适配，这样声音模型和相关的参数适合于新的词句、用户，并且如果必要的话，适合于周围的噪声和有关的麦克风位置。在这里，新入口被加到含有由声音模型表示的语音识别设备词汇的语音识别设备词典，或者已经存在的词汇入口被适配。在使用隐藏的马尔可夫模型（HMM）的语音识别设备中，隐藏的马尔可夫模型的蜕变概率和输出分布被调整。

在训练期间，语音识别设备适应于一个还没有操作过该设备的用户的语音输入。这能由一个用户在该语音识别设备的第一次使用之前或者甚至更迟的在语音输入未被触发的第一次使用之后发生。利用以前的语音数据实行重新训练也是有可能的。训练性能的不同变量是已知的，其中，例如，分布混合的估计是在固定的 HMM 拓扑结构中或者同样该数量的一个变更或者 HMM 状态的转换中发生的。

在语音识别设备的一个适配中，语音识别设备的预置参数被适配并被最佳化，其中在由一个用户完成的训练中该参数被预置或者以已经执行的适配处

与完成的语音输入相关连的语音识别结果为基础来训练或适应该语音识别设备。

5 依靠根据本发明的的方法训练或适配不需要任何训练或适配模式的启动。甚至在训练或适应使用之前用户输入确定也不是必需的。用于该训练或适配的语音输入和相关连的语音识别结果的适用性宁可通过以语音识别结果为基础的实际情况建立，一个可指配的设备功能通过例如运行语音拨号处理被激活，并且如果用户不给出一个具有他表示他的拒绝的输入，例如停止设备功能的执行（权利要求2），则它假设语音识别结果表示正确的语音输入并能被用作训练和适配。例如，用户也能通过语音或键盘输入表示其拒绝。

10 权利要求3涉及一种在语音识别结果被用作训练或适配之前附加的安全措施，对于这个结果，倘若这里用户没有表示其拒绝，即使该语音识别结果不表示正确的语音输入，则一个训练操作或适配不以这个语音识别结果为基础而被执行。这个安全措施被设计为防止用户恰好被分散注意力并且由此不能表示其拒绝的情况，即使该语音识别结果是不能令人满意的，依据一个设备功能的想要的行为也不发生。

15 权利要求4涉及在权利要求3中描述的安全措施防止以不令人满意的可靠性级别为基础的训练和适配的情况。如果，直接在语音输入之后通过例如键盘输入的另一输入方法完成一个输入，这里前面的语音输入和通过其他输入方法完成的输入的一个高概率将发生。为了估计其一致性，考虑到检测没有发生的该语音输入和下面通过其他方法完成的输入的情况的一个可靠性级别同样被建立。如果该可靠性级别落在一个预定的目标范围内，一个训练操作和适配被考虑语音识别结果和通过其他输入方法完成的输入执行。

20 权利要求5，象权利要求4一样，涉及一个语音输入最初没有被考虑适合于训练或适配，不过接下来的输入是通过另一种输入方法完成的以便再次有一个通过其他方法完成的输入和前面的语音输入的同时发生的较高概率的情况。对照权利要求4，权利要求5结合在一个语音输入之后，参照这个结果用户使一个输入表示其拒绝，出于这个原因，最初没有训练或适配发生的情况。这里在通过其他输入方法完成该输入之后的训练和适配之前，利用考虑到检测没有发生的该语音输入和下面通过其他方法完成的输入的情况的一个可靠性

级别实施评估。

在如权利要求 6 阐述的实施例 5 中，对于每个语音识别结果至少一个可替换的语音识别结果被产生。如果，以最后的语音输入和相应的语音识别结果为基础，一个训练操作和适应被执行，如果最后之前的语音输入没有导致任何训练或适配，它同样检查是否该语音识别结果对应于一个对于最后之前的语音输入的可变更的语音识别结果的最后语音输入。如果是这种情况，它假定最后之前的语音输入和最后的语音输入彼此对应。在随后的额外训练操作或随后的额外适配中，语音识别设备的参数被以这样一种方式设定，最后之前的语音输入和最后的语音输入都将在新的输入操作中导致正确的语音识别结果，也就是导致对于最后的语音输入是确定的的语音识别结果。

在如权利要求 7 阐述的实施例 15 中，一个用户在听觉上或视觉上被通知一个语音识别结果。因此，在语音识别结果的准确度评定中，这个用户不仅仅参考一个设备功能的性能的识别。当设备功能的性能对于该用户是不可见的，或者他仅仅被给出关于由该语音识别设备产生的语音识别结果的不完全信息的时候，用户需要额外的信息。

权利要求 8 和 9 涉及本发明在一个移动电话中的应用。在这里，该语音识别设备特别地用于语音拨号。

权利要求 10 涉及一种用于实现根据本发明的方法的电子设备。在权利要求 11 中，该电子设备采用移动电话终端的形式。

20 附图说明

参考附图中所示的实施例的例子本发明将被做进一步描述，然而，本发明不受此限制。

图 1 示出了一个移动电话终端的方框电路图；和

图 2 至 5 示出了本发明各种变型的流程图。

25 具体实施方式

图 1 示出了一个作为一个具有一个数字信号处理器 2 的移动电话终端形成的电子设备 1，其首先被装配用于语音编码 / 解码和免提信号处理（被一个功能块 3 结合）并其次用于一个语音识别设备的语音识别程序的运行（功能块 4）。该移动电话终端 1 具有一个麦克风 5 和一个扬声器 6，它们通过模数转换器 7

和8连接到该数字信号处理器2。该移动电话终端1还含有一个微控制器9用于移动电话终端1的功能的中心控制。此外,一个存储设备10被提供,其用于语音识别设备4的词典11和其它数据12的存储,在这里,为了简化,其它数据由功能块12结合——因为它们对于本发明不是必需的。词典11以作为隐藏的马尔可夫模型产生的声音模型的形式含有语音识别设备4的词汇,词典11必须被认作语音识别设备4的一部分。移动电话终端1也包括一个键盘输入单元13和一个显示单元14。通过一个天线15,高频信号被接收并发射通过收发信机单元16处理的高频信号。一个信道编码/解码单元17被用于将被发送的信号的信道编码或接收信号的信道解码。一个总线系统18连接数字信号处理器2,微控制器9,存储设备10,键盘输入单元13,显示单元14和信道编码/解码单元17。

依靠语音识别设备4通过借助于麦克风5给出的语音输入实现移动电话终端1的功能是可能的。在上面举例的情况中,通过一个语音输入,另一个用户的拨叫被触发(语音拨号)。在这个过程中,一个用户说想要被呼叫的其他用户的任意一个电话号码,或者他说其他用户的名字,那些人的关联电话号码必须已经预先分配给这个名字,对于这种分配任意一种单独的语音输入或借助于键盘输入单元13的单独的键盘输入,或者键盘输入和语音输入二者能被使用。为了保证语音识别设备4的平稳运行,语音识别设备4的训练是必要的,也就是一个词汇适合于应用在问题中并存储在必须建立的词典11中。同样,考虑到使该语音识别模块适应变更的应用环境,稍后该已经被训练的语音识别设备4的声音模块的适配被提供。

图2至图5示出了用于该语音识别设备4的训练或适配的各种实施例。

在图2所示的配置中,首先在步骤200语音输入模式被激活,例如通过按下一个相应的启动按键。在一个步骤201,用户进行一个语音输入以便他输入他想要呼叫的另一个用户的名字。在步骤202,在语音识别设备4的帮助下,执行一个自动语音识别操作并产生一个语音识别结果。在步骤203这个语音识别结果被输出给用户,例如通过借助于显示单元14的视觉输出或者依靠一个借助于扬声器6的声音输出。在步骤204,假定一个电话号码已经被分配给这个已识别的名字,这个电话号码的自动拨号被立刻处理。在步骤205,进行一

个关于该自动拨号是否被用户中断(例如在用户识别出一个错误的语音识别结果并拒绝它的情况下,通过按下一个特定按键或讲一个“停止”语音指令)的询问,进行一个返回到步骤 200 和步骤 201 之间的点的跳转,在那里该设备准备接收一个进一步的语音输入。如果该语音拨号没有被中断,并且在步骤 205
5 给这个询问一个否定的答案,以步骤 202 确定的该语音识别结果为基础执行语音识别设备 4 的训练和适配。这样对于一个进一步的语音输入能被进行的结果,进行一个返回到步骤 200 和步骤 201 之间的点的跳转。

在根据图 3 的实施例中,最初在一个步骤 300 移动电话终端 1 的语音输入模式被启动。在此之后,在步骤 301 完成一个语音输入,也就是在提问的情况下,一个名字被讲。在步骤 302,在一个语音识别操作已经被执行之后,在
10 步骤 303 对于给用户听觉上或视觉上该语音识别结果被重新产生。这样,如在权利要求 2 中已经被示出的,具有该识别的名字的一个用户的自动拨叫被触发。在步骤 305,进行一个关于该自动拨号是否被用户中断的询问。如果是这种情况,进行一个返回到步骤 300 和步骤 301 之间的点的跳转,因为它已经假定该用户以已经认为该语音识别结果为错误的并拒绝它。如果自动拨号没有被
15 中断,在步骤 306,建立一个该语音识别结果的可靠性级别评定。在步骤 307,进行一个关于是否该可靠性级别落在一个预定的目标范围内的询问。如果不是这种情况,进行一个返回到步骤 300 和步骤 301 之间的点的跳转。如果该可靠性级别落在指定的目标范围内,它必须假定该语音识别结果实际上正确地表示了语音输入并且在步骤 308 以这个语音识别结果为基础执行训练操作或适
20 配。

在根据图 4 的实施例中,首先在一个步骤 400 语音输入模式被启动。接着在步骤 401 完成一个语音输入,也就是讲他想要呼叫的一个用户的名字。在步骤 402,执行关于该已经被完成的语音输入的一个语音识别操作。在步骤 403
25 该语音识别结果在声音上或视觉上被输出给用户。在步骤 404,具有已识别名字的一个用户的自动拨叫被触发。在步骤 405,执行一个关于该拨叫是否被用户中断的检测。如果这是肯定情况,也就是说如果该语音识别结果被用户识别为错误,则进行一个返回到步骤 400 和步骤 401 之间的点的跳转。如果自动拨号没有被用户中断,对于该语音识别结果在步骤 406 确定一个初始的可靠性级

5 别。在步骤 407 中，建立一个关于是否这个初始的可靠性级别落在一个最初指定的目标范围内的询问。如果这是肯定情况，在步骤 408，以这个语音识别结果为基础执行语音识别设备 4 的训练或适配。如果在步骤 407 这个初始的可靠性级别没有落在一个最初指定的目标范围内被建立，在步骤 409 执行一个关于是否通过其它输入方法，例如通过键盘输入完成随后输入的检测。如果不是这种情况，再次进行一个返回到步骤 400 和步骤 401 之间的点的跳转。然而如果在步骤 409，通过其它输入方法—也就是特别地通过键盘输入—完成下面以后的直接的输入被建立，早先的语音输入和其后的通过其它输入方法的输入相互一致的可能性是很大的，也就是在一个失败的语音拨号之后，该用户尝试通过借助于另一种输入方法的输入触发对同一个用户的拨叫。下面通过其它输入方法输入，在步骤 410 考虑到借助于另一种输入方法完成的输入，对于前面的语音识别结果一个第二可靠性级别被确定。在步骤 411，执行一个关于是否这个第二可靠性级别落在一个第二指定的目标范围内的检测。如果这是肯定情况，象步骤 408 一样，考虑到借助于其它输入方法完成的输入，以这个语音识别结果为基础执行该语音识别设备的训练或适配。如果在步骤 411，这个第二可靠性级别没有落在一个第二指定的目标范围内被建立，这与依靠其它输入方法完成的输入和前面的语音输入彼此不一致的验证一致，并完成一个返回到步骤 400 和步骤 401 之间的点的跳转。

20 原则上，即使在步骤 405 该用户中断一个用户的拨叫并由此表示该语音识别结果的拒绝被建立的情况下，代替一个返回到步骤 401 和 402 之间的点的跳转，继续进行步骤 409 的询问是可能的，如一个由虚线表示的跳转 412 指示的。因此，如果其后一个输入被立即通过另一种输入方法完成，它再次被假定具有一个通过其它输入方法完成的输入和前面的语音输入彼此对应的高的概率级别，并且该语音识别结果不能表示正确的语音输入。然而，即使该语音识别结果不是完全正确的，如果这里对该语音输入和/或通过另一种输入方法完成的输入有一个足够的近似性，它也能够被用于该语音识别设备的训练或适配，它通过基于一个可靠性级别执行的检测被建立。该训练操作因此既包括包含在语音识别结果中的信息也包括包含在通过其它输入方法完成的输入中的信息。

最后，图 5 示出了一个本发明的变化，其中应用有区别的训练或适配。在步骤 500，该语音输入模式被启动，在步骤 501 完成一个语音输入，也就是讲一个用户的名字，并且在步骤 502 通过语音识别设备 4 执行一个语音识别操作，其中除该语音识别结果之外的至少一个可替换的进一步的语音识别结果被产生。在步骤 503，该语音识别结果在声音上或视觉上被输出到用户。在步骤 504，关于该语音识别结果，执行具有已识别名字的一个用户的自动拨叫。在步骤 505，执行一个关于该拨叫是否被用户中断的检测。如果这是肯定情况，进行一个返回到步骤 500 和步骤 501 之间的点的跳转。如果这不是肯定情况，在步骤 506，以这个最后的语音识别结果为基础执行语音识别设备 4 的训练或适配。在一个询问步骤 507，一个关于是否这个最后的语音识别结果对应于一个由最后之前的语音输入确定的可替换的语音识别结果，和是否该最后之前的语音输入导致了被用户自动拨叫的停止的进一步的检测接着被执行。如果这些条件之一不被满足，进行一个返回到步骤 500 和步骤 501 之间的点的跳转。如果这两个条件都被满足，它假定所述对最后之前的语音输入的可替换的语音识别实际上表示了最后之前的语音输入的正确语音识别结果。假如这样的情况下，在步骤 508 执行有区别的训练或有区别的适配操作，提供最后的语音输入和一个最后之前的表示相同的词语或相同的词语序列。与在词典 11 输入的相应的词汇相关联的该声音模型以这样一种方式适应，对于最后的语音输入和对于最后之前的一个语音输入语音识别设备 4 都将产生正确的语音识别结果。在步骤 508 被执行之后，再次进行一个返回到步骤 500 和 501 之间的跳转。

本发明即不限制于应用在一个移动电话终端中，也不限制于语音拨号。依靠语音输入一个动作能被在电子设备的另一种功能上完成，例如，在语音输出的声音设定上。同样，对于一个用户进行的语音识别的听觉或视觉输出来说，尤其是如果该用户能够从对应于一个语音输入的一个设备的特性，而没有任何进一步的信息中推断出语音识别结果的准确度，它不是绝对必需的。此外，该用户也能表示其拒绝，另外或换句话说，通过不中断已经启动的一个设备功能的预置输入（例如于置语音或键盘输入）。

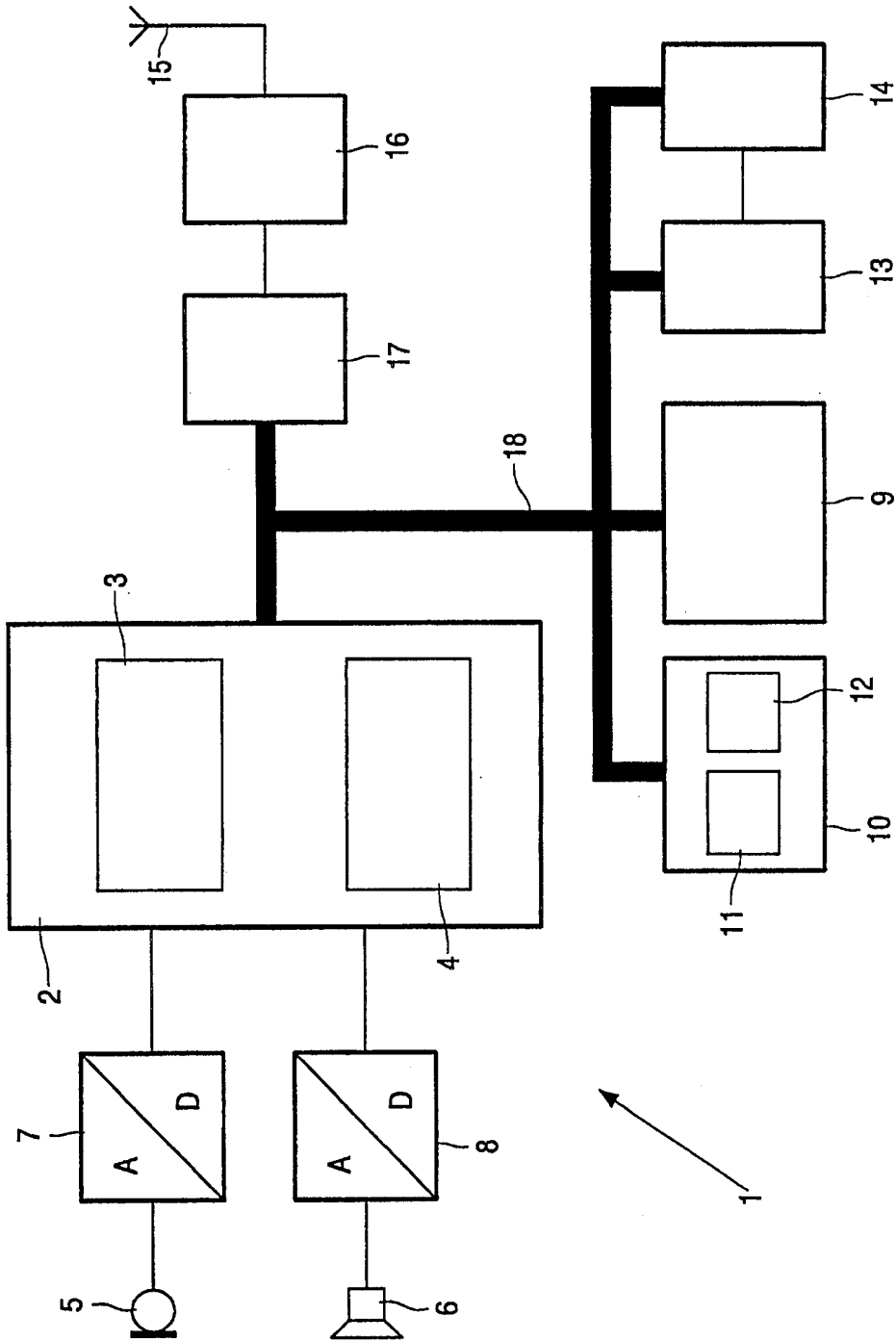


图 1

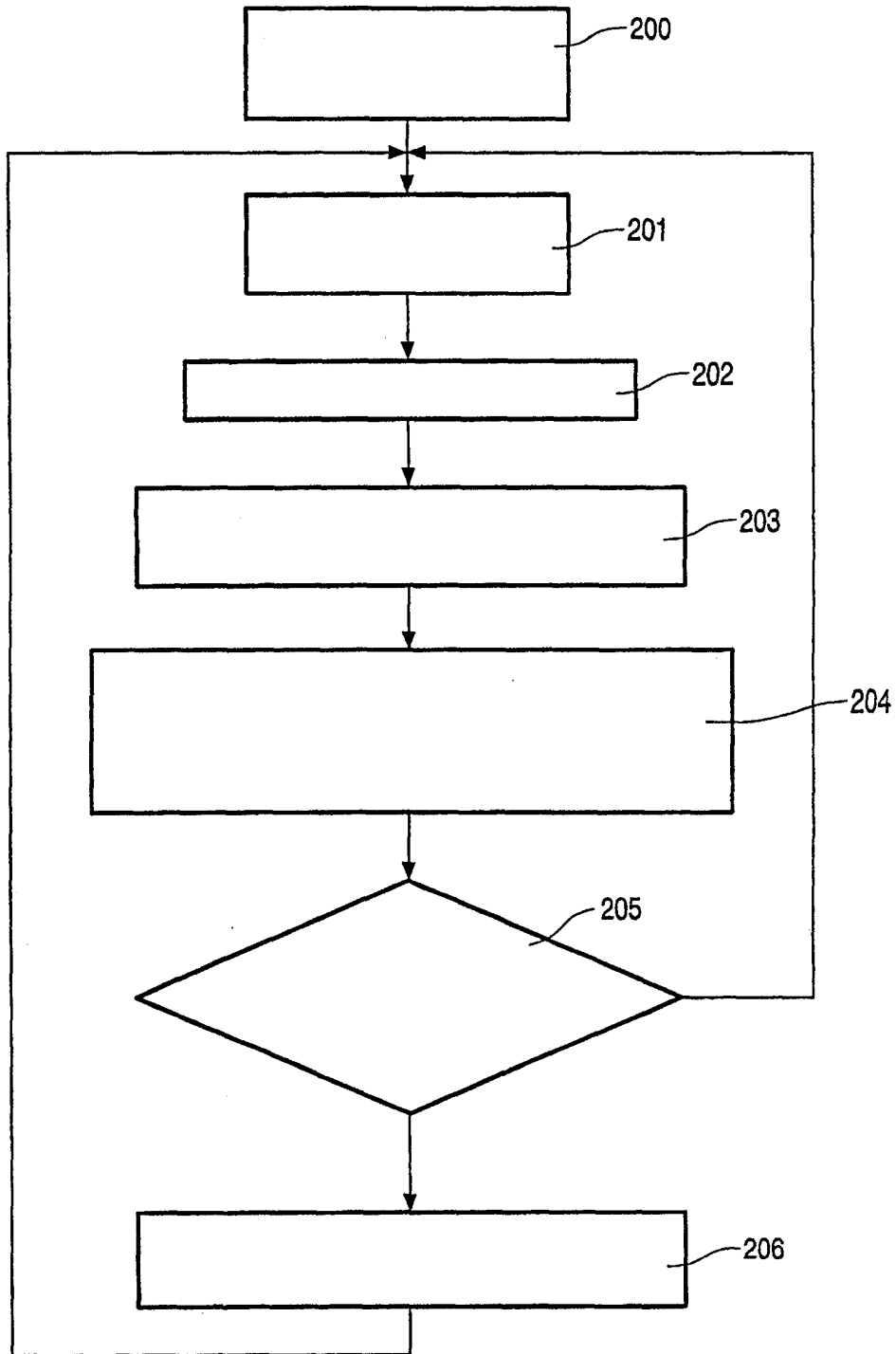


图 2

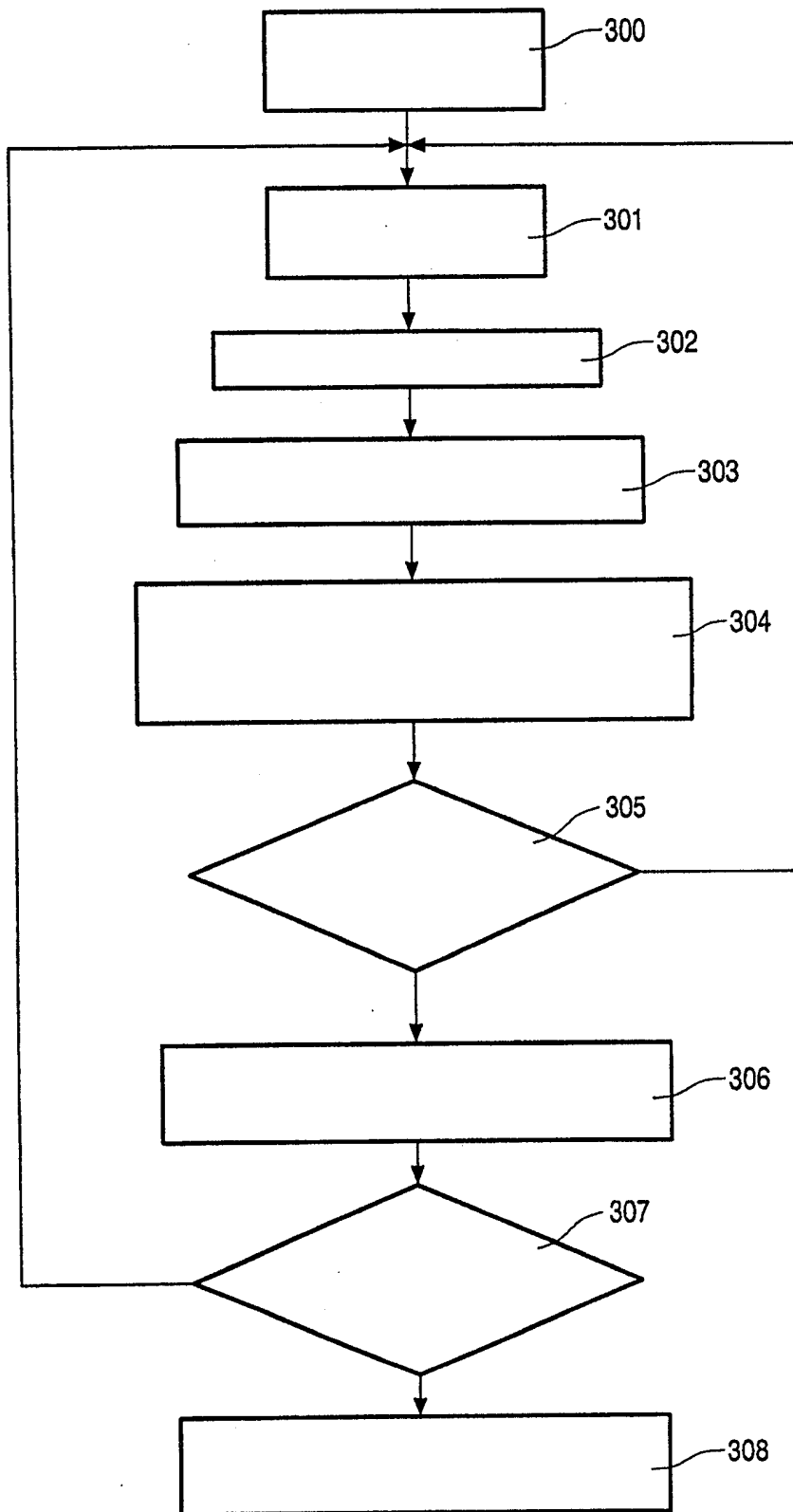


图 3

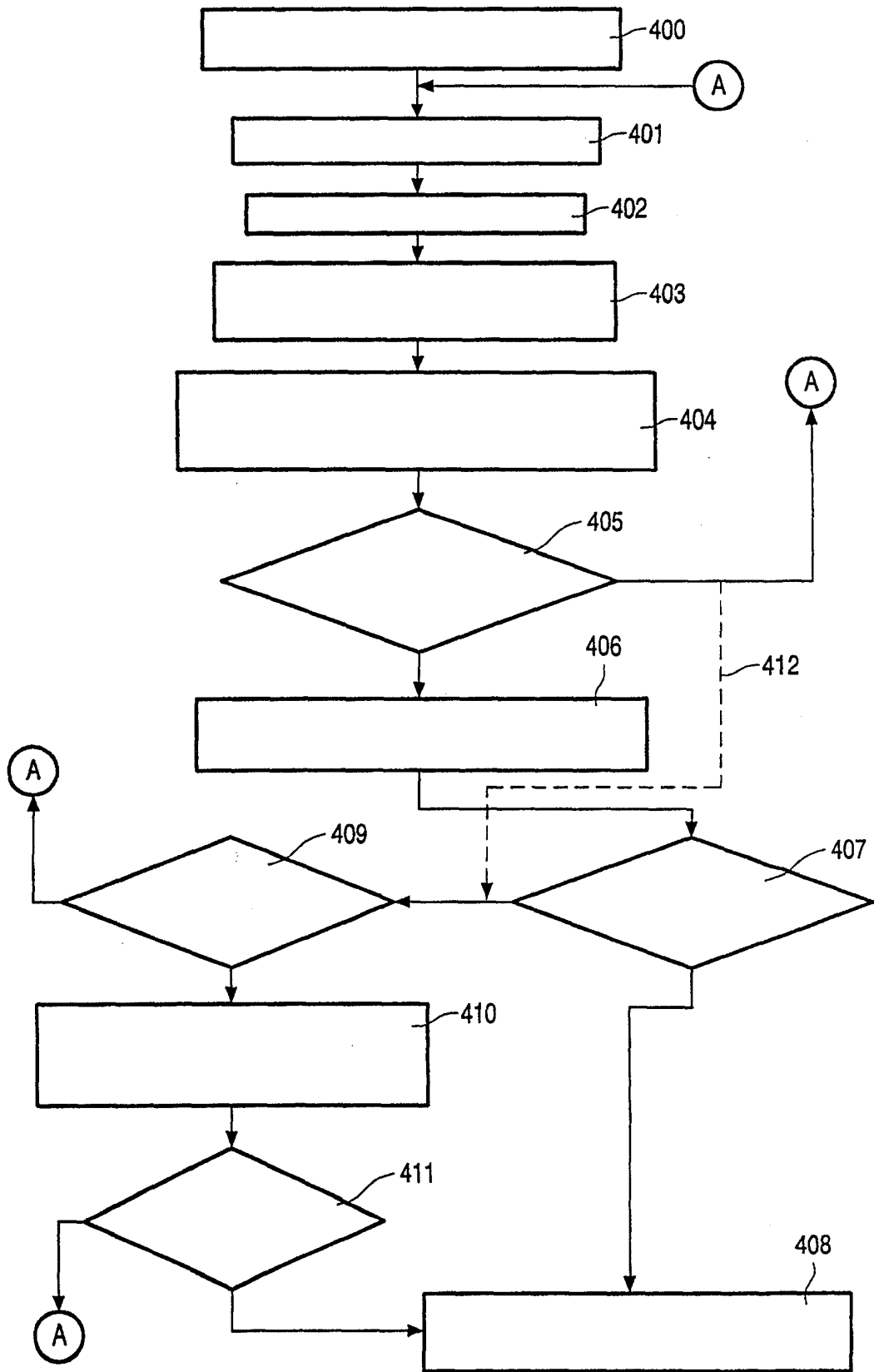


图 4

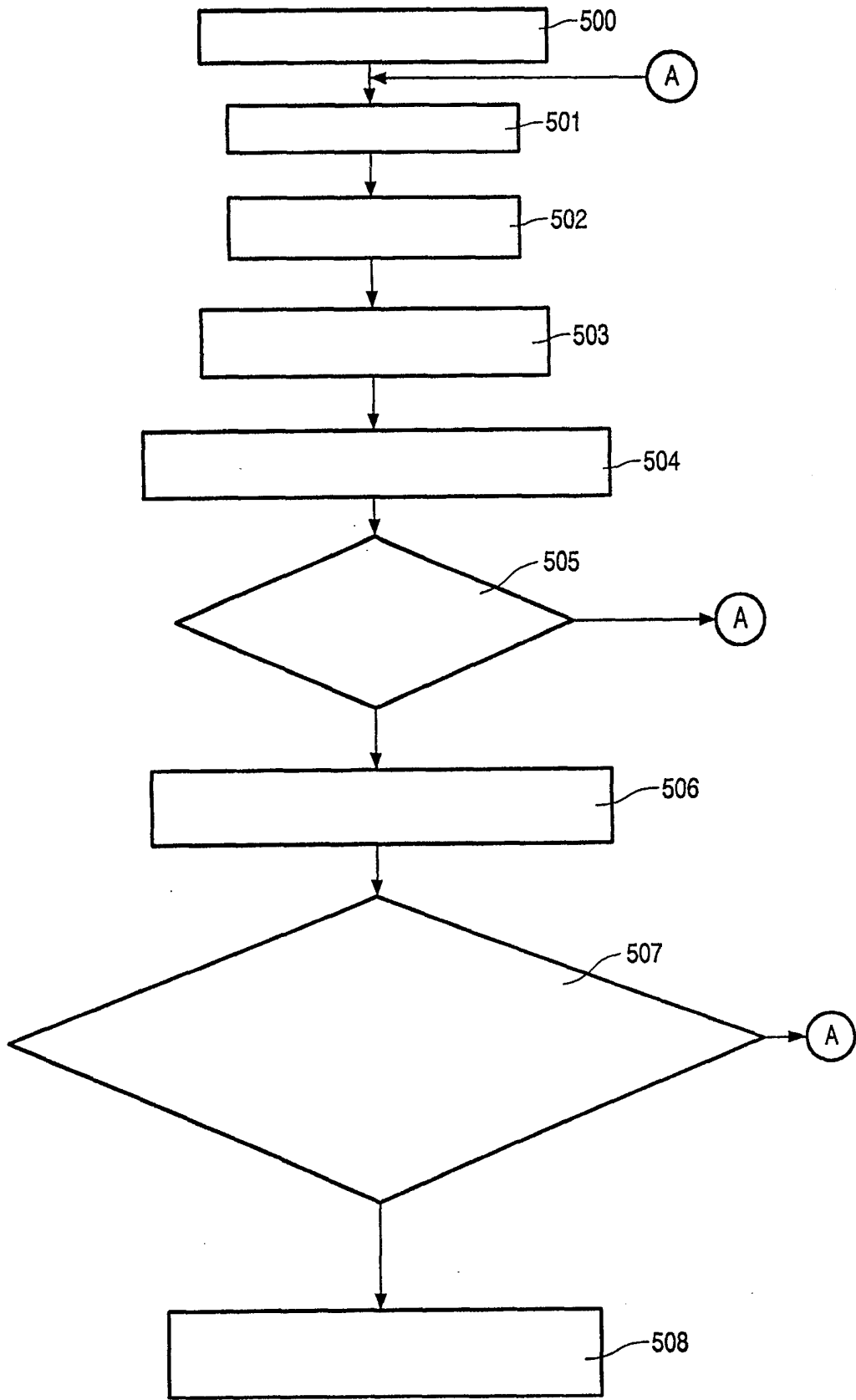


图 5