

## (12) 发明专利申请

(10) 申请公布号 CN 102576529 A

(43) 申请公布日 2012.07.11

(21) 申请号 201080045971.8

(74) 专利代理机构 中科专利商标代理有限责任

(22) 申请日 2010.09.08

公司 11021

## (30) 优先权数据

2870/CHE/2009 2009.11.20 IN  
12/652,608 2010.01.05 US

## (41) 代理人 戎志敏

(85) PCT申请进入国家阶段日

## (51) Int. Cl.

2012.04.12

G10L 15/20 (2006.01)

G10L 21/02 (2006.01)

## (86) PCT申请的申请数据

PCT/IB2010/054050 2010.09.08

## (87) PCT申请的公布数据

W02011/061636 EN 2011.05.26

(71) 申请人 印度科学院

地址 印度卡纳塔克邦

(72) 发明人 尼尚斯·乌哈斯·奈尔

塞坡·梵卡塔那拉沙亚·斯里尼瓦

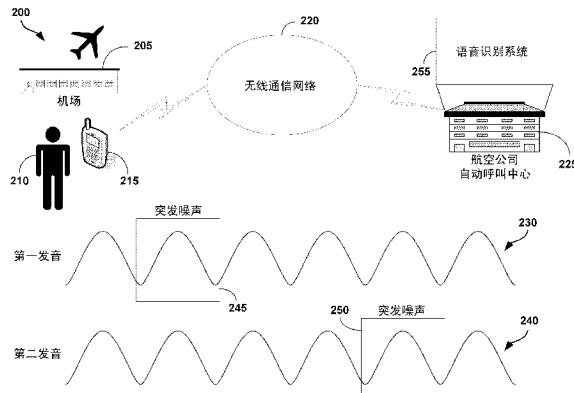
权利要求书 2 页 说明书 17 页 附图 8 页

## (54) 发明名称

针对多个模式的联合解码使用多模式维特比  
算法的系统和方法

## (57) 摘要

公开了使用多模式维特比算法对多个模式进行联合解码的系统、设备和方法。示例性方法可以接收针对 K 个信号重复中的每一个的多个时序信号观测值集。进一步，每个信号观测值集与具有时间索引点的 K 维时间网格的相应维度相关联。此外，在多个时间索引点中的每个时间索引点处，利用处理器计算针对隐式马尔科夫模型 (HMM) 状态集中的每个状态的状态成本度量。此外，针对状态集中的每个状态，针对给定的时间索引点，状态成本度量计算提供了最可能前驱状态和相应的最可能前驱时间索引点。该示例性方法还可以使用所计算的状态成本度量确定状态序列，以及确定针对 HMM 的相应累积概率测量。



1. 一种方法，包括：

接收针对 K 个信号重复中的每一个信号重复的多个时序信号观测值集，其中每个信号观测值集与具有时间索引点的 K 维时间网格的相应维度相关联；

在多个时间索引点中的每个时间索引点处，利用处理器计算针对隐式马尔科夫模型 HMM 状态集中的每个状态的状态成本度量，其中针对状态集中的每个状态，针对给定的时间索引点，状态成本度量计算提供了最可能前驱状态和相应的最可能前驱时间索引点；

使用所计算的状态成本度量来确定状态序列，以及确定针对 HMM 的相应累积概率测量。

2. 根据权利要求 1 所述的方法，还包括：

针对多个时间索引点，重复计算状态集中的每个状态的状态成本度量，以及确定针对多个 HMM 的最可能序列和相应累积概率测量。

3. 根据权利要求 1 或 2 所述的方法，其中针对多个时间索引点中的给定时间索引点，通过以下步骤来确定状态集中的每个状态的状态成本度量：

在每个可能的前驱时间索引点处，计算与每个可能的先前状态相关联的成本度量；以及

选择每个状态的最低成本度量。

4. 根据权利要求 3 所述的方法，其中针对给定的可能前驱状态，状态成本度量仅基于与在从给定前驱时间索引点移到多个时间索引点中的给定时间索引点时递增的维度相关联的观测值。

5. 根据前述任一权利要求所述的方法，其中确定最可能状态序列包括：在终端时间索引点处，标识最终状态处的最低状态成本度量。

6. 根据前述任一权利要求所述的方法，其中信号观测值是信号特征矢量。

7. 根据权利要求 6 所述的方法，其中特征矢量是从由以下参量构成的组中选择的：梅尔倒谱系数特征矢量、线性预测编码系数、谱密度、谱能量、噪声比、声长、相对功率、和滤波器匹配。

8. 根据前述任一权利要求所述的方法，其中多个时间索引点是关于与穿过 K 维空间的对角线相距的预定距离来确定的。

9. 根据权利要求 8 所述的方法，其中每个观测值集具有相应的持续时间，以及预定距离基于相应的持续时间之差。

10. 根据前述任一权利要求所述的方法，其中使用所计算的状态成本度量来确定状态序列包括：基于状态成本度量来回溯经过时间索引点。

11. 根据前述任一权利要求所述的方法，其中计算针对给定状态的给定状态成本度量基于：与所有候选前驱时间索引点相关联的所有状态的状态成本度量、从每个候选前驱时间索引点的每个状态转变到给定状态的概率、从相应候选前驱时间索引点转变的相应概率、以及从状态集中的状态发射的观测值的联合概率。

12. 根据前述任一权利要求所述的方法，其中所确定的状态序列确定观测值集的校准。

13. 一种制造产品，包括其上存储有指令的计算机可读介质，如果由计算设备执行所述指令，则使计算设备执行以下操作：

从存储器中获取针对 K 个信号重复中的每一个的 K 个时间序列信号观测值集，其中每

个信号观测值集与具有时间索引点的 K 维时间网格的相应维度相关联；

从存储器中获取针对多个隐式马尔可夫模型 HMM 中的每一个 HMM 的参数集；

在多个时间索引点中的每个时间索引点处，针对给定 HMM 状态集中的每个状态来计算状态成本度量，其中，针对状态集中的每个状态，针对给定的时间索引点，状态成本度量计算提供最可能前驱状态和相应的最可能前驱时间索引点；

确定针对多个 HMM 中的每一个 HMM 的累积概率测量；以及

确定多个 HMM 中的最可能 HMM。

14. 一种设备，包括：

执行存储在存储器中的软件应用的处理器，软件指令包括：

在多个时间索引点的每个时间索引点处，计算针对给定 HMM 状态集中的每个状态的状态成本度量，其中，针对状态集中的每个状态，针对给定的时间索引点，状态成本度量计算提供了最可能前驱状态和相应的最可能前驱时间索引点；

使用所计算的状态成本度量来确定状态序列，以及确定针对多个 HMM 中的每一个 HMM 的相应的累积概率测量；以及确定多个 HMM 中的最可能 HMM。

15. 根据权利要求 14 所述的设备，还包括：存储器，用于存储：

针对 K 个信号重复中的每一个的、多个时间序列信号观测值集的数字表示，其中每个数字观测值集与具有时间索引点的 k 维时间网格的相应维度相关联；以及

针对多个隐式马尔可夫模型 HMM 中的每一个 HMM 的参数集。

16. 根据权利要求 14 或 15 所述的设备，还包括：音频接收机，用于：

接收针对 K 个信号重复中的每一个的多个时序音频信号观测值集，其中，每个音频信号观测值集与具有时间索引点的 k 维时间网格的相应维度相关联；以及

将针对 K 个信号重复中的每一个的多个时序音频信号观测值集转换为针对 K 个信号重复中的每一个的多个时序模拟电信号观测值集。

17. 根据权利要求 14 至 16 之一所述的设备，还包括：模拟到数字转换器，用于将针对 K 个信号重复中的每一个的多个时序模拟电信号观测值集转换为针对 K 个信号重复中的每一个的多个时序信号观测值集的数字表示。

18. 根据权利要求 14 至 17 之一所述的设备，其中处理器基于针对 K 个信号重复的每一个的多个时序信号观测值集的数字表示来计算多个特征矢量集。

19. 根据权利要求 14 至 18 之一所述的设备，其中处理器使用所计算的状态成本度量将状态序列转换为状态序列的数字表示。

20. 根据权利要求 14 至 19 之一所述的设备，还包括数字到模拟转换器，用于将状态序列的数字表示转换为输出模拟电信号。

## 针对多个模式的联合解码使用多模式维特比算法的系统和方法

### 背景技术

[0001] 在每天的电话 / 移动会话中,会话中的收听者通常会由于他们无法理解说话者语音的特定部分而请说话者重复语音的特定部分。这种情形在有严重影响语音可辨度的背景噪声的情况下更为常见。尤其在有重 / 突发背景噪声的情况下,语音识别系统、设备和方法可以利用这种重复的信息,以针对各种应用,更好地辨别语音。

[0002] 一些语音识别系统,如自动语音识别 (ASR) 系统,在测试和训练条件可兼容时工作良好。ASR 系统的示例可以是用于航空公司的自动呼叫中心的语音识别系统。许多语音识别系统 (包括 ASR 系统) 存储包括表示最可能使用的语音部分的数据的训练数据。训练数据不受周围噪声、不同的说话者口音、或对于语音数据的任何其他负面音频效果的影响。然而,真实测试环境不同于训练条件。诸如加性噪声、回声和说话者口音之类的各种因素会影响许多真实测试环境中的语音识别性能。由于 ASR 可以表征为统计模式识别问题,如果测试模式不同于用于训练模型的任何模式,则会出现错误。已经提出了各种增强 ASR 技术中的鲁棒性的方式,包括:(i) 降低模型的可变性,或者(ii) 修改统计模型参数以适合噪声环境。然而,在极高噪环境或突发错误信道的情况下,如在可能丢弃分组的分组通信中,语音识别系统可能会受益于采用使用重复发音来准确地解码语音的方案。

### 发明内容

[0003] 本申请公开了使用多模式维特比算法检测多个模式下的信号的系统、设备和方法。

[0004] 本公开的一个实施例可以是一种接收针对 K 个信号重复中的每一个的多个时序信号观测值集的方法。进一步,每个信号观测值集与具有时间索引点的 K 维时间网格的相应维度相关联。此外,在多个时间索引点中的每个时间索引点处,利用处理器计算针对隐式马尔科夫模型 (HMM) 状态集中的每个状态的状态成本度量。此外,针对状态集中的每个状态,针对给定的时间索引点,状态成本度量计算提供了最可能前驱 (predecessor) 状态和相应的最可能前驱时间索引点。在一个实施例中,该方法包括:确定最终状态的成本度量和终端时间索引点。这也被称为通过相应的 HMM 生成观测值的累积概率测量。因而,一些方法还包括:针对多个 HMM 中的每一个 HMM,在终端时间索引点处确定最终状态的成本度量,然后选择最小成本度量和相应的 HMM。然后,相应的 HMM 用于标识该模式(例如,该模式可以是语音识别系统中的词)。一些方法还可以包括:使用所计算的状态成本度量来确定状态序列,以及确定该 HMM 的相应累积概率测量。

[0005] 此外,一些方法包括:针对多个时间索引点,重复计算状态集中的每个状态的状态成本度量,以及针对多个 HMM,确定最可能序列和相应的累积概率测量。之后,该方法可以针对多个 HMM,基于相应的累积概率测量或成本度量来标识最可能 HMM。该方法还可以包括:针对多个时间索引点中的给定时间索引点,通过在每个可能的前驱时间索引点处,计算与每个可能的先前状态相关联的成本度量,以及选择每个状态的最低成本度量,来确定状态

集中的每个状态的状态成本度量。

[0006] 在一些实施例中,针对给定的可能前驱状态,状态成本度量仅基于与在从给定前驱时间索引点移到多个时间索引点中的给定时间索引点时递增的维度相关联的观测值。此外,一些方法可以包括:通过在终端时间索引点处,标识最终状态处的最低状态成本度量,来确定最可能状态序列。可以使用多个时间索引点,以限制所使用的时间维网格中的点,并且可以关于与穿过 K 维空间的对角线相距的预定距离来确定多个时间索引点。预定距离可以基于观测值序列的各个持续时间之差。

[0007] 该方法可以包括:基于与所有候选前驱时间索引点相关联的所有状态的状态成本度量、从每个候选前驱时间索引点的每个状态到给定状态转变的概率、从相应候选前驱时间索引点转变的相应概率、以及从状态集中的状态发射的观测值的联合概率,来计算给定状态的给定状态成本度量。所确定的状态序列还可以确定观测值集的校准(alignment)。

[0008] 这里所描述的一些实施例可以采用产品形式,包括其上存储有指令的计算机可读介质,如固态存储器、压缩盘、数字视频盘 ROM、磁存储介质等,如果由计算设备执行该指令,则使计算设备执行包括以下的操作:从存储器中获取针对 K 个信号重复中的每一个的 K 个时间序列信号观测值集,其中每个信号观测值集与具有时间索引点的 K 维时间网格的相应维度相关联;从存储器中获取针对多个 HMM 中的每一个 HMM 的参数集;在多个时间索引点中的每个时间索引点处,针对给定 HMM 状态集中的每个状态来计算状态成本度量,其中,针对状态集中的每个状态,针对给定的时间索引点,状态成本度量计算提供了最可能前驱状态和相应的最可能前驱时间索引点;确定针对多个 HMM 中的每一个 HMM 的累积概率测量;以及确定多个 HMM 中的最可能 HMM。

[0009] 在其他实施例中,该设备包括:执行存储在存储器中的软件应用的处理器,软件指令包括:在多个时间索引点的每个时间索引点处,计算针对给定 HMM 状态集中的每个状态的状态成本度量,其中,针对状态集中的每个状态,针对给定的时间索引点,状态成本度量计算提供了最可能前驱状态和相应的最可能前驱时间索引点;可选地,使用所计算的状态成本度量来确定状态序列,确定针对多个 HMM 中的每一个 HMM 的相应的累积概率测量;以及确定多个 HMM 中的最可能 HMM。

[0010] 在一些实施例中,该设备还包括:存储器,用于存储针对 K 个信号重复中的每一个的、多个时间序列信号观测值集的数字表示,其中每个数字观测值集与具有时间索引点的 k 维时间网格的相应维度相关联;以及用于存储针对多个 HMM 中的每一个 HMM 的参数集。

[0011] 在其他实施例中,该设备还包括:音频接收机,用于接收针对 K 个信号重复中的每一个的多个时序音频信号观测值集,其中,每个音频信号观测值集与具有时间索引点的 k 维时间网格的相应维度相关联;以及用于将针对 K 个信号重复中的每一个的多个时序音频信号观测值集转换为针对 K 个信号重复中的每一个的多个时序模拟电信号观测值集。

[0012] 在其他实施例中,该设备还包括:模拟到数字转换器,用于将针对 K 个信号重复中的每一个的多个时序模拟电信号观测值集转换为针对 K 个信号重复中的每一个的多个时序信号观测值集的数字表示。

[0013] 以上概述仅是示例性的,并不意在进行任何限定。除了上述示意性的方面、实施例和特征之外,其他方面、实施例和特征也将通过参考附图和以下的具体实施方式而显而易见。

## 附图说明

- [0014] 图 1 是示例语音识别系统的隐式马尔可夫模型的示例。
- [0015] 图 2 示出了包括多模式维特比算法的方案的示例性语音识别系统。
- [0016] 图 3 是使用多模式维特比算法的示例语音识别系统的功能框图。
- [0017] 图 4 是描述了使用多模式维特比算法来解码由语音识别系统接收的多语音发音中的语音的示例方法的示例流程图。
- [0018] 图 5 是在多模式维特比算法中使用  $K = 2$  个语音模式来解码语音的示例时间路径。
- [0019] 图 6 是示出了在多模式维特比算法中使用  $K = 2$  个语音模式的最优状态序列和最优时间路径的三维网格。
- [0020] 图 7 是示出了被设置用于使用多模式维特比算法的语音识别系统的示例计算设备 700 的方框图。
- [0021] 图 8 是使用多模式维特比算法从多个信号观测值中检测信号的示例方法 800 的流程图。

## 具体实施方式

- [0022] 在以下详细说明中,参考了作为详细说明的一部分的附图。在附图中,类似符号通常表示类似部件,除非上下文另行指明。具体实施方式部分、附图和权利要求书中记载的示例性实施例并不是限制性的。在不脱离在此所呈现主题的精神或范围的情况下,可以利用其他实施例,且可以进行其他改变。应当理解,在此一般性记载以及附图中图示的本公开的各方案可以按照在此明确和隐含公开的多种不同配置来设置、替换、组合、分割和设计。
- [0023] 这里所描述的是针对多模式的联合解码使用多模式维特比算法的系统、设备和方法。
- [0024] 通常,这里所描述的实施例包括使用多模式维特比算法 (MPVA) 的鲁棒语音识别技术。考虑到与人类电话通信的类似性,由于收听者无法理解说话者,收听者可以请说话者重复他们语音的特定部分。这种情形在有严重影响语音可辨度的背景噪声的情况下更为常见。在极高噪环境或突发错误信道的情况下,如在可能丢弃分组的分组通信中,语音识别系统可能会受益于采用在实现语音识别技术的过程中使用重复发音的方案。尽管 MPVA 可用于语音识别和方法,MPVA 还可以用于从多个模式中检测信号的任何系统和方法。
- [0025] 此外,可以在包括移动电话技术、命令和控制应用、火车站中的语音识别、军事应用、机器人技术、发音估计及许多非语音应用的各种应用中使用这些实施例。许多应用可能需要准确辨识在有显著的不利背景噪声的情况下的说话者的语音。例如,移动电话中的语音识别系统在有像车辆噪声、马路噪声等的瞬态噪声的情况下并不会工作良好。实施例可以允许移动电话用户重复用户想要呼叫的人的名字,从而尤其在有重 / 突发噪声的情况下提高语音识别性能。其他实施例可以包括在诸如飞行员想要给出指令的嘈杂的驾驶舱之类的命令和控制应用中。类似地,实施例可以用于诸如在其中许多人在背景中说话(被称为串音噪声)的火车站之类的嘈杂环境中。
- [0026] 在这种军事应用中,士兵可以与包括语音识别系统的自动化设备进行通信。因而,

语音识别系统对在由于来自机枪的子弹、来自大炮的炮壳等而导致的高度环境噪声的战场上的士兵的语音进行解码。此外，实施例可以用于机器人工业应用，其中机器人可以在工厂或其他工业环境中使用来自人类控制者的语音的多次重复来学习 / 识别命令。其他实施例可以应用于对来自多个模式的发音进行联合估计的发音估计。实施例还可以包括在广泛使用的各种应用，如语音识别、生物信息学、远程通信、语言学、图像处理、关键词定位等，以及可以使用动态编程（例如，维特比算法）的任何应用。

[0027] 隐式马尔可夫模型 (HMM) 和动态编程可以用于许多语音识别技术。HMM 是建模的系统被假设为具有未观测状态的马尔可夫过程的统计模型。马尔可夫过程是无记忆系统的数学模型，它在任何给定时刻的给定未来状态的似然概率仅取决于其当前状态，而不取决于任何过去状态。在常规马尔可夫过程中，状态对于观测者是直接可见的，因而状态变换概率是唯一参数。

[0028] 马尔可夫过程的示例可以是掷硬币的结果序列。掷硬币的结果可以被建模为随机变量。针对掷硬币的随机变量，存在两个同样可能的结果，正和反各具有 0.5 的概率。随机变量的所有结果的概率之和为 1。此外，随机过程是随机变量的序列，因而，掷硬币的结果序列可以被建模为随机过程。此外，掷硬币的结果并不取决于先前掷硬币的结果，因而可以被描述为无记忆的。因此，掷硬币的结果序列可以被建模为马尔可夫过程。

[0029] 不是无记忆的随机过程的示例可以是从袋中无置换地拾取彩色弹珠的结果。例如，袋内可以包含五个黑弹珠和五个白弹珠。从袋中的第一拾取为黑弹珠的概率是 0.5。然而，从袋中的第二拾取为黑弹珠的概率取决于第一拾取的结果。如果第一拾取为黑弹珠，则第二拾取为黑弹珠的概率为  $4/9 = 0.44$ 。相反，如果从袋中的第一拾取为白弹珠，则第二拾取为黑弹珠的概率为  $5/9 = 0.56$ 。因此，拾取弹珠的特定结果的概率取决于过去的结果。因而，这种随机过程不是无记忆的。

[0030] 不同于马尔可夫过程，在隐式马尔可夫模型中，状态不是直接可见，而是取决于状态观测输出事件。每个状态具有保持该状态或变换为另一状态的变换概率。此外，每个状态针对每个输出事件具有发射概率。因此，由 HMM 生成的输出事件序列给出了关于状态序列的一些信息。术语“隐式”指模型经过的状态序列，并不是指模型的参数（如，变换概率或发射概率）。即使确切地知道模型参数，但因为状态对于观测者来说并不可见，该模型仍然是“隐式”的。隐式马尔可夫模型尤其以它们在时间模式识别中的应用（如语音、手写、手势识别、词性标注、乐谱跟踪、局部放电和生物信息学）而广为人知。

[0031] 图 1 是示例语音识别系统的隐式马尔可夫模型 100 的示例。HMM100 可以对例如航空公司的自动呼叫中心的潜在呼叫者的语音发音进行建模。示例语音识别系统可以用于辨别呼叫者发音中的航班号。此外，航空公司可以将所说的词“Five”作为航班号中的数字。因而，用于语音识别系统中的 HMM 可以包括“Five”的音素作为状态。音素是用于形成发音间有意义对比的声音的最小分割单元。例如，“F”、“i”、“ve”是词“Five”中包含的三个不同音素。因而，图 1 示出了 HMM 的一部分，显示了表示词“Five”中的前两个音素“F”和“i”(105, 120) 的两个状态。此外，HMM 具有观测到的输出“F”和“i”(160, 135)。HMM 100 还示出了从一个状态变换到下一个状态 (115, 170, 180) 或保持在该状态 (110, 115) 的状态变换概率，以及从每个状态到观测到的输出 (140, 145, 150 和 155) 的发射概率。发射概率是在给定 HMM 100 的状态下观测事件的似然概率。所有变换到某一状态的变换概率 (110,

170, 115, 125) 之和以及从某一状态变换出去的变换概率 (110, 115, 125, 180) 之和等于 1。此外, 发射到观测到的输出的发射概率之和也等于 1 (140, 145, 150, 155)。图 1 中的 HMM 100 可以仅是由语音识别系统使用的更大 HMM 的一部分。因而, 可以有如由概率为 0.3 的变换 170 所指示的从不同于“i”的状态到状态“F”的状态变换。此外, 可以有如由概率为 0.3 的变换 180 所指示的从“i”到不同于“F”的状态的状态变换。

[0032] 通常, 当实施应用时, 找出 HMM 中的最可能状态序列是有用的。许多不同方法可以用于找出 HMM 中的最可能状态序列。动态编程是通过以递归方式将复杂问题分为较简单的子问题来简化复杂问题的数学优化方法, 并且可以用于找出 HMM 中的最可能状态序列。此外, 维特比算法是用于找出产生隐式马尔可夫模型中的观测事件序列的最可能隐式状态序列 (称为维特比路径) 的示例动态编程算法。

[0033] 这里所描述的实施例利用多模式维特比算法 (MPVA), 这是可以用于许多解码和信号检测应用 (可以分析多个模式) 的新型动态编程方法。这种应用之一是语音识别系统。在示例性语音识别系统中, 对多个发音进行联合解码, 以识别语音模式。本领域技术人员将会理解, MPVA 还可以用于可以使用动态编程方法的任何应用。

[0034] 图 2 示出了使用 MPVA 的示例性语音识别应用 200。呼叫者 210 可以通过无线通信网络 220, 使用移动电话 215 与航空公司自动呼叫中心 225 联系并通信。自动呼叫中心可以具有语音识别系统 255, 用于接收来自航空公司顾客 (如呼叫者 210) 的呼叫。此外, 例如, 语音识别系统可以向呼叫者 210 请求航班号, 以为呼叫者 210 访问航班信息。然后, 呼叫者 210 可以说出航班号的数字, 如“Five”。语音识别系统 255 可以请求呼叫者 210 重复发音“Five”。可以通过图 2 中的图形表示 (230, 240) 所示的音频信号来表示航班号的数字“Five”的每个发音。

[0035] 在两个发音的图形表示 (230, 240) 中, 示出了具有多于一个发音的语音识别系统 255 对呼叫者 210 给出的航班号进行解码的益处。由于呼叫者 210 的机场周围 205 的环境噪声, 突发噪声会影响每个发音 (230, 240) 中音频信号的不同部分 (245, 250)。因此, 通过对单个发音的音频信号进行处理, 语音识别系统 255 会对呼叫者 210 所说的航班号进行不准确地解码。然而, 通过处理包含相同音素序列 (例如, 航班号) 的两个发音 (230, 240), 语音识别系统 255 可以接收两个音频信号, 其中突发噪声影响了每个信号的不同部分 (245, 250)。因此, 语音识别系统可以使用 MPVA, 利用两个重复发音, 对呼叫者 210 所说的语音进行准确解码。

[0036] 语音识别系统可以使用 HMM 来帮助从一个或多个接收到的音频信号中解码语音。通过语音识别系统接收到的每个接收音频信号可以采用许多不同形式之一。处理后的信号可以被分析为用于语音识别系统中的 HMM 的观测到的输出的时间序列, 或简单地, 观测值的时间序列。音频信号的一个示例性处理可以是针对音频信号计算梅尔倒谱系数 (MFCC)。在一些实施例中, 可以针对每 20 毫秒的采样音频数据来计算 MFCC 矢量。在一些实施例中, 可以使用重叠间隔来计算 MFCC, 如通过对 20 毫秒的音频数据进行处理, 然后平移 10 毫秒, 并对该间隔中的音频数据进行处理, 依次类推。可以通过对音频信号的分贝谱进行处理 (如采用傅立叶变换) 来找出倒谱系数。MFCC 可以是语音识别系统中音频信号的许多不同特征或观察值之一。特征是声音现象的单独可测量启发式特性, 可以作为语音识别系统中 HMM 的观测到的输出。特征或观测值可以包括音频信号的一部分的 MFCC、谱密度、谱能量、

噪声比、声长、相对功率、滤波器匹配等。

[0037] 图3是使用多模式维特比算法解码器的示例语音识别系统的功能框图。示例语音识别系统300可以用于航空公司自动呼叫中心。自动呼叫中心可以向呼叫者请求航班号，以访问所请求的信息。此外，自动呼叫中心可以请求呼叫者多次重复航班号，以确保对呼叫者语音的准确识别。

[0038] 当呼叫者说出航班号时，如“Five”或“Nine”，语音识别通过使用接收机310来接收表示呼叫者的发音的音频信号。接收机310可以是麦克风、声换能器、或将音频信号转换为模拟电信号的一些其他音频接收机。接收机可以将模拟电信号转发至模拟到数字转换器(ADC)315，以将模拟电信号变换为表示该模拟电信号（进而该音频信号）的数字数据。模拟到数字转换器315可以将数字数据存储在系统存储器320的信号存储部分330中。当然，可以通过多种方式来提供采样音频数据：可以通过手机和/或公共交换电话网络的一部分、或者通过与计算机工作站相关联的麦克风和ADC来提供接收机和ADC，因而它们并不是系统的必要组件。

[0039] 此外，处理器350可以是示例语音识别系统的一部分。处理器350可以包含MFCC子处理器360，该MFCC子处理器360可以访问和处理所存储的音频信号的数字数据，以获得梅尔倒谱系数(MFCC)，作为语音识别系统的特征矢量。然后，词或词组的给定发音的特征矢量时间序列将形成该给定发音的观测值的时间序列。

[0040] 处理器还可以包括MPVA解码器370，该MPVA解码器370接收MFCC特征矢量形式的观测值，并访问系统存储器320的HMM存储部分340中的HMM数据。HMM数据包括通常表示为 $\lambda$ 的已知参数。

[0041] MPVA解码器370执行MPVA，以对从存储器设备320获取的多个发音中的语音进行解码。此外，存储器设备320可以存储可以控制处理器350上MPVA的执行的程序指令。

[0042] 在一些实施例中，系统可以可选地包括数字到模拟转换器(DAC)375。DAC375将表示解码语音的数字数据变换为模拟电信号。此外，DAC将模拟电信号转发至系统输出接口380，该系统输出接口380可以是将模拟电信号转换为表示解码语音的音频信号的扬声器或一些其他声换能器设备。因此，语音识别系统300可以向呼叫者叙述音频信号，以便呼叫者验证解码语音。如果呼叫者指示解码语音不准确，则语音识别系统可以请求呼叫者清楚地表达另一重复发音，使得MPVA可以有更多数据来对呼叫者的语音进行准确解码。

[0043] 图4是描述了使用多模式维特比算法来对由语音识别系统接收到的多个语音发音中的语音进行解码的示例方法的示例流程图400。解码语音的预备性步骤是选择要用于语音识别的音频信号的特征405。语音发音的特征的示例可以是语音发音的梅尔倒谱系数。另一预备性步骤可以是训练包括观测到的输出的多个状态、状态变换概率和发射概率密度函数的一个或多个HMM410。可以通过分析无噪语音来开发HMM。此外，可以训练HMM，以对具有不同口音(美语、英语等)的、有男有女的多个不同说话者的语音发音“Voice Dialer”进行分析。这种语音发音可以包括8个状态，每个状态对应于语音发音中的音素，“V”“oi”“ce”“D”“i”“a”“l”“er”。在训练这种HMM期间，可以找出语音发音“Voice Dialer”的状态变换和发射概率。在一个实施例中，在系统中使用多个HMM：针对要标识的每个模式都有一个单独的HMM。在语音识别系统中，每个HMM对应于不同的词(或词组)。系统可以通过训练导出HMM，或可以将HMM提供给系统。

[0044] 实施示例方法 400 的语音识别系统可以接收表示来自呼叫者 415 的重复语音发音的多个音频信号。多个音频信号可以由音频接收机、麦克风、声换能器或一些其他设备接收。示例方法 400 中的另一步骤可以是将每个音频信号处理为每个音频信号的数字数据表示 425。处理可以包括：通过音频接收机、麦克风、声换能器等将每个音频信号变换为模拟电信号。之后，可以使用模拟到数字转换器，将每个模拟电信号转换为数字数据。数字数据可以存储在语音识别系统的存储器中。在其他实施例中，可以向系统提供数字化音频数据采样，并存储于存储器中。模拟到数字转换和获得采样音频数据的相关方式和方法对于本领域技术人员来说是已知的，并不重要。

[0045] 此外，语音识别系统可以基于表示与发音相对应的每个音频信号的数字数据来计算梅尔倒谱系数 (MFCC) 430。处理器或子处理器可以访问系统存储器中的数字数据，以计算 MFCC。MFCC 是要由 MPVA 使用的观测值的时间序列。此外，MPVA 可以使用除 MFCC 之外的不同特征，如 LPCC、声长、噪声比等。LPCC 表示使用线性预测模型信息的压缩形式的数字语音信号的谱包络。

[0046] 在另一步骤 440 中，语音识别使用系统存储器中存储的一个或多个 HMM，将 MPVA 应用于 MFCC。可以通过在执行软件指令的处理器中具体化的 MPVA 解码器来实现 MPVA。当应用 MPVA 时，语音识别系统接收分别具有帧长  $T_1, T_2, \dots, T_K$  的、表示为  $O_{1:T_1}^1, O_{1:T_2}^2, \dots, O_{1:T_K}^K$  的  $K$  个观测值序列，其中  $O_{1:T_i}^i = (O_{1:T_i}^i, O_{2:T_i}^i, \dots, O_{T_i:T_i}^i)$  和  $O_{t_i}^i$  是时间帧  $t_i$  处第  $i$  模式的特征矢量集。 $K$  个观测值序列中的每一个可以属于相同的模式类别（例如，所说的词），因而通过单个 HMM 进行解码，表示为  $\lambda$ 。然而，语音识别系统可以使用各对应于词的一个或多个 HMM。此外， $K$  个观察值序列可以是同一说话者相同词的不同发音。MPVA 能够对这  $K$  个模式进行联合解码，以辨别说话者的语音。

[0047] 使用 MPVA 对多个接收到的发音进行联合解码，为了使其直观化，考虑标准 HMM 格式结构到  $K+1$  维的扩展，其中  $K$  维对应于  $K$  个模式，因而定义了  $K$  维时间网格，一维对应于给定 HMM 的 HMM 状态集。与标准 HMM 的二维格式结构相类似，格式结构网格遍历沿所有  $K$  个时间轴  $T_1$  到  $T_K$  单调。这是因为，重复发音的观测值以前向方式联合地经过时间。可以看出，如果精确地同步（例如，在相同的相对时间偏移处出现所有音素）重复模式（例如，针对语音识别算法的同一词的多个发音），则将会期望最佳路径是沿穿过  $K$  维时间索引空间的对角线。另一方面，如果发音并未彼此完美地记录，如所期待地，模式的相应部分的  $K$  个观测值将不会沿对角线排列。事实上，针对每个模式或发音，观测值序列典型具有不同长度。因此，如下所述，最优路径可以不沿对角线，算法提供了每个观测值集中的观测值的不同使用率或消耗（因为每个观察值集与  $K$  个时间维之一相关联）。通过 HMM 状态变换矩阵来确定沿  $(K+1)$  维的状态变换（遍历的或从左至右的）。

[0048] HMM 状态序列可以定义如下：

$$[0049] \quad \mathbf{q} = \mathbf{q}_{\phi(1):\phi(T)} \stackrel{\text{def}}{=} [\mathbf{q}_{\phi(1)}, \dots, \mathbf{q}_{\phi(T)}] \quad (1) \text{ 其中 } q_{\phi(i)} \in 1:N \text{ 是时间}$$

$\phi(t) = (t_1, \dots, t_K)$  处的状态索引，其中  $\phi(t)$  通过  $K$  维网格中的点来表示， $N$  是 HMM 状态的个数。由于状态序列取决于  $\phi(t)$  的演进 (evolution)，可以在  $K$  维时间空间或网格上定义另一变量，如  $\Phi = [\Phi(1), \Phi(2), \dots, \Phi(T)]$  (见图 5)，使得  $\Phi(1) = (1, \dots, 1)$ ，

$\Phi(T) = (T_1, \dots, T_K)$ 。 $\Phi$  是  $K$  个模式之间的时间校准。可以通过  $\{q_\phi(t), \phi(t)\}$  来表示  $(K+1)$  维格式结构网格中的任一点。此外,  $t$  表示沿  $K$  维网格 (见图 5) 中的每个坐标的跳数, 以及  $\phi(t) = (t_1, \dots, t_K)$  表示  $K$  维空间中的单个点。此外,  $t$  从 1 移至  $T$ , 其中  $T$  是从  $(1, \dots, 1)$  到  $(T_1, \dots, T_K)$  所需的总跳数。 $T$  的值取决于所经过的路径, 最优时间路径的  $T$  值可以仅在路径回溯之后获知。

[0050] MPVA 的三个目标可以是确定下式 (2)–(6) 中所示的总联合多模式似然、最优状态序列和最优时间路径。最优状态序列展现了在语音识别系统中使用 MPVA 时的解码语音。

[0051] 总联合多模式似然可以定义如下:

$$[0052] P(\mathbf{O}_{1:T_1}^1, \dots, \mathbf{O}_{1:T_K}^K; \lambda) = \sum_{\forall q} P(\mathbf{O}_{1:T_1}^1, \dots, \mathbf{O}_{1:T_K}^K, \mathbf{q}_{\phi(1):\phi(T)}; \lambda) \quad (2)$$

[0053] 考虑通过  $(K+1)$  维网格的所有有效路径, 沿最优 HMM 状态序列  $q^*$  和最优时间路径  $\Phi^*$  的联合  $K$  模式似然可以定义如下:

$$[0054] P(\mathbf{O}_{1:T_1}^1, \dots, \mathbf{O}_{1:T_K}^K, \mathbf{q}^*, \Phi^*; \lambda) = \max_{\forall (\mathbf{q}, \Phi)} P(\mathbf{O}_{1:T_1}^1, \dots, \mathbf{O}_{1:T_K}^K, \mathbf{q}, \Phi; \lambda) \quad (3)$$

[0055] 式 (2) 是关于给定 HMM  $\lambda$  的所有  $K$  个模式的总概率。此外, 可以如下找到最大似然 (最优) HMM 状态序列  $q^*$  和最优时间路径  $\Phi^*$ :

$$[0056] (\mathbf{q}^*, \Phi^*) = \arg \max_{(\mathbf{q}, \Phi)} P(\mathbf{q}, \Phi / \mathbf{O}_{1:T_1}^1, \dots, \mathbf{O}_{1:T_K}^K; \lambda) \quad (4)$$

$$[0057] (\mathbf{q}^*, \Phi^*) = \arg \max_{(\mathbf{q}, \Phi)} P(\mathbf{q}, \Phi, \mathbf{O}_{1:T_1}^1, \dots, \mathbf{O}_{1:T_K}^K; \lambda) \quad (5)$$

$$[0058] (\mathbf{q}^*, \Phi^*) = \arg \max_{(\mathbf{q}, \Phi)} P(\mathbf{O}_{1:T_1}^1, \dots, \mathbf{O}_{1:T_K}^K / \mathbf{q}, \Phi; \lambda) P(\mathbf{q}, \Phi; \lambda) \quad (6)$$

[0059] 通过遍历  $(K+1)$  维网格来联合地确定  $(\Phi^*, q^*)$ 。在网格中, 在单个步骤中, MPVA 以宽度第一方式从  $\Phi(1)$  移到  $\Phi(T)$ , 覆盖全部时间轴。可以计算通过网格的部分路径的递归更新 (与标准 HMM 维特比算法相类似)。项  $\delta_{\phi(t)}(j)$  可以定义如下:

[0060]

$$\delta_{\phi(t)}(j) = \max_{\{q_{\phi(1)}, \dots, q_{\phi(t-1)} = j, \phi(1), \dots, \phi(t-1)\}} P(\mathbf{O}_{1:t_1}^1, \dots, \mathbf{O}_{1:t_K}^K, q_{\phi(1)}, \dots, q_{\phi(t)} = j, \phi(1), \dots, \phi(t); \lambda) \quad (7)$$

[0061]  $\delta_{\phi(t)}(j)$  是在经过多维网格时的累积似然, 可以描述为从状态  $i$  到  $j$  及从时间索引点  $\phi(t-1)$  到  $\phi(t)$  的最低成本度量。经过网格意味着,  $K$  模式的一部分与 HMM 状态相匹配。因而,  $\delta_{\phi(t)}(j)$  可以被认为是部分模式似然的测量或成本度量。每个模式是长度为  $T_1$  或  $T_2$  或  $\dots$   $T_K$  的 MFCC 矢量序列。MPVA 的应用可以描述为具有多个步骤。这些步骤可以包括但不限于初始化 445、递归 450、终止 455 和路径回溯 460。

[0062] 在初始化步骤 445 中, 初始概率  $\delta_{\phi(1)}(i)$  可以表示为:

$$[0063] \delta_{\phi(1)}(i) = P(O_1^1, \dots, O_1^K, q_{\phi(1)} = i, \phi(1); \lambda) = \pi_i P(\phi(1)) b_i(O_1^1, \dots, O_1^K) \quad (8)$$

[0064] 其中  $i = 1, \dots, N$ ;  $\pi_i = P(q_{\phi(i)} = i)$  是初始状态分布;  $b_i(O_1^1, \dots, O_1^K) = P(O_1^1, \dots, O_1^K / q_{\phi(1)} = i, \lambda)$ 。因此,  $\delta_{\phi(1)}(i)$  是式 (7) 的初始值, 并且可以被认为是最佳路径 (例如,  $\{1, 1, 1, \dots, 1\}$ ) 起点处的似然测量, 使用来自所有重复发音的第一观测值集。在每个所允许的 HMM 状态中分配起始概率, 并且由  $\pi_i$  来控制该起始概率。此外,  $b_i(O_1^1, \dots, O_1^K)$  是 HMM 状态  $i$  的概率密度函数。

[0065] 在递归步骤 450 中,  $\Delta\phi(t) = \phi(t) - \phi(t-1) = (\Delta t_1^1, \Delta t_2^2, \dots, \Delta t_K^K)$ , 从而  $0 \leq \Delta t_i^i \leq 1$ , 至少一个  $\Delta t_i^i$  具有非零值。矢量中非零值的位置提供了关于在从  $\phi(t-1)$  移到  $\phi(t)$  的过程中递增或遍历哪些维的指示。 $\Delta\phi(t)$  可以包括至少一个非零值, 最多 K 个非零值。可以用于 MPVA 中的另一约束可以包括限制在时间上的后向移动。也就是说, 当如下所述向时间索引 K 维网格填充状态成本度量  $\delta_{\phi(t)}$  时, 可以删除特定的可能前驱时间索引点  $\phi(t-1)$  不考虑。这些约束形成遍历多维网格的局部连续性约束 (LCC) 集。针对 K 个模式, 对于每个  $\phi(t)$ , 示例性 LCC 可以是存在  $(2K-1)$  个可能  $\phi(t-1)$ 。也可以选择其他类型的 LCC。

[0066] 此外,  $S_{\phi(t)} = \{O_{t_i}^i \mid \Delta t_i^i \neq 0, i = 1, 2, \dots, K\}$  可以是在  $\phi(t)$  处一起进行映射的观测值集, 如 MFCC 矢量。此外,  $\{O_{\phi(t)}\} = \{O_{t_m}^m, \dots, O_{t_n}^n\}$ , 从而  $\{O_{t_m}^m, \dots, O_{t_n}^n\}$  是集  $S_{\phi(t)}$  中的所有特征矢量。此外,  $\{0_{\phi(t)}\}$  可以是矢量  $(O_{t_1}^1, O_{t_2}^2, \dots, O_{t_k}^k)$  的子集, 仅保留  $\Delta t_k^k$  非零的那些  $O_{t_k}^k$ 。集  $S_{\phi(t)}$  和  $\{0_{\phi(t)}\}$  可以具有最少一个特征矢量, 最多 K 个特征矢量。为零的  $\Delta t_k^k$  可以指示: 由于信号中的噪声、时间规整等, 不发射该时间索引  $O_{t_k}^k$  处的特征矢量。因而, 可以示出如下:

$$[0067] \quad \delta_{\phi(t)}(j) = \max_{\{q_{\phi(t-1)}=i, \phi(t-1)\}} \{\delta_{\phi(t-1)}(i) a_{ij} P(\phi(t) / \phi(t-1)) b_j(\{O_{\phi(t)}\}) \quad (9)$$

[0068]  $\phi(t)$  从  $\phi(1) = (1, \dots, 1)$  改变到  $\phi(T) = (T_1, \dots, T_K)$ ;  $i, j = 1, 2, \dots, N$ 。此外,  $a_{ij}$  是从状态  $i$  到状态  $j$  的状态变换概率 (如在标准 HMM 中),  $P(\phi(t) / \phi(t-1))$  是从  $\phi(t-1)$  移到  $\phi(t)$  的概率, 以及  $b_j(\{O_{\phi(t)}\})$  是由状态  $j$  发射的  $\{O_{\phi(t)}\}$  的联合似然。此外,  $b_j(\{O_{\phi(t)}\})$  与由状态  $j$  发射的所有矢量  $\{O_{t_m}^m, \dots, O_{t_n}^n\}$  的联合似然相同, 其中  $\{O_{t_m}^m, \dots, O_{t_n}^n\}$  包括集  $S_{\phi(t)}$  中的所有特征矢量。因此, HMM 状态  $j$  可以从 K 个模式中发射可变数量的矢量, 该可变数量对应于  $\Delta\phi(t)$  矢量中的非零值的个数。但是, 无论采用了哪个时间路径  $\Phi$ , 当  $\delta_{\phi(t)}$  的递归计算到达  $\delta_{\phi(T)}$ , 每个状态  $j$  将会发射明确数量的多模式特征矢量 =  $(T_1 + T_2 + \dots + T_K)$ 。

[0069] 换言之, 式 (9) 的一部分的一个解释是: 在 K 维网格中的每个时间索引点  $\phi(t)$  处, 确定 HMM 中每个状态的状态成本度量。这通过查看每个可能前驱点  $\phi(t-1)$  处与每个状态  $i$  相关联的成本度量  $\delta_{\phi(t-1)}(i)$ 、以及  $i$  到  $j$  状态变换概率  $a_{ij}$  和从每个候选前驱点  $\phi(t-1)$  移到  $\phi(t)$  的变换概率来执行。更具体地, 可以通过计算与每个可能前驱时间索引点处的每个可能先前状态相关联的成本度量, 然后选择每个状态的最低成本度量, 来计算网格中给定点的状态成本度量。状态成本度量计算提供最可能前驱状态, 还提供相应的最可能前驱时间索引点。

[0070] 式 9 描述了最大概率, 然而, 本领域技术人员将会理解, 式 9 将会通过对式 9 的自变量的对数取反, 并以“min”函数替换“max”函数, 来描述成本度量。

[0071] 递归步骤向 K 维网格中的每个点或至少点的子集填充  $\delta_{\phi(t)}(j)$  值。通过从左到右、从上至下遍历网格来进行该填充。针对每个被填充的点, 通过选择有效前驱点中的最小成本选择, 从前驱点到当前点递增成本度量; 可以在系统存储器中存储从每个网格点的各种候选前驱点中选取的“最佳前驱”的选择, 以在之后执行路径回溯。因此, MPVA 可以直到完成执行并到达网格的终点 (即,  $\{T_1, T_2, \dots, T_K\}$ ), 才标识出最优状态序列或最优时间路径, 然后执行路径回溯步骤。图 5 示出了在多模式维特比算法 500 中使用 K = 2 个模式 ( $O_{1,T_1}^1$  (轴 t1) 和  $O_{1,T_2}^2$  (轴 t2)) 的语音来解码语音而产生的示例时间路径  $\Phi$ 。在图 5 中, 如果坐标 (5, 5) 被认为是  $\phi(t)$ , 则根据路径  $\Phi$ ,  $\phi(t-1)$  是 (5, 4)。可以有许多这样的可能

路径  $\Phi$ , MPVA 需要从中选择最优路径  $\Phi^*$ 。在时刻 (3,3) 处, 通过状态 j 发射特征矢量  $O_3^1$  和  $O_3^2$ 。在时刻 (4,3) 处, 因为已经使用了  $O_3^2$ , 所以仅发射矢量  $O_4^1$ 。发射可变数量的矢量, 不存在矢量的再用。一方面, 状态成本度量仅基于与从给定前驱时间索引点  $\phi(t-1)$  移到给定时间索引点  $\phi(t)$  时递增的维相关联的观测值。

[0072] 图 5 还示出了当  $K = 2$  时使用的 LCC 580。也就是说, 针对二维网格中的每个点  $\phi(t) = (t_1, t_2)$ , 存在  $2K-1$  个前驱点。因此, 当  $\phi(t) = (t_1, t_2)$  时,  $\phi(t-1)$  可以是  $(t_1-1, t_2)$ 、 $(t_1, t_2-1)$  或  $(t_1-1, t_2-1)$ 。如前所述, 可以删除特定候选前驱点不考虑, 以降低计算的复杂度。类似地, 如下所述, 可以利用其他限制或简化, 包括全局约束。

[0073] 此外, 状态和时间维的回溯指针可以表示为  $\Psi_{\phi(t)}(j)$  和  $\Gamma_{\phi(t)}(j)$ , 可以通过下式获得:

$$[\Psi_{\phi(t)}(j)\Gamma_{\phi(t)}(j)] = \arg \max_{\{q_{\phi(t-1)}=i, \phi(t-1)\}} \{\delta_{\phi(t-1)}(i)a_{ij}P(\phi(t)/\phi(t-1))b_j(\{O_{\phi(t)}\}) \quad (10)$$

[0075] 其中,  $\phi(t)$  从  $(1, \dots, 1)$  改变到  $(T_1, \dots, T_K)$ ;  $i, j = 1, 2, \dots, N$ 。在计算  $P(\phi(t)/\phi(t-1))$  的过程中, 如果  $\phi(1) = (1, 1, \dots, 1)$ , 则  $P(\phi(1)) = 1$ , 否则,  $P(\phi(1)) = 0$ 。

[0076] 可以有多种可能方式来计算  $P(\phi(t)/\phi(t-1))$ 。为了从  $\phi(t-1)$  到达  $\phi(t)$ , MPVA 可以已经定义了 LCC。以下提出了两种可能方法, 可以开发其他特定 (ad hoc) 方法。

[0077] 在一个实施例中, 假设从一个时刻移到另一个时刻的均匀概率分布, 可以得到以下概率:

$$[0078] P(\phi(t)/\phi(t-1)) = \frac{1}{L} \quad (11)$$

[0079] 其中 L 是从其可以到达  $\phi(t)$  的  $\phi(t-1)$  的可能位置的总数。可以通过 LCC 来定义  $\phi(t-1)$  的可能值的范围。通常  $L = 2^K - 1$ , 除非  $\phi(t)$  位于多维网格的边界。

[0080] 在另一实施例中, MPVA 可以赋予 LCC 的对角线区域更大权重, 如下:

$$[0081] P(\phi(t)/\phi(t-1)) = \frac{\|\phi(t) - \phi(t-1)\|}{\sum_{\forall \phi(t-1)} \|\phi(t) - \phi(t-1)\|} \quad (12)$$

[0082] 其中  $\|\cdot\|$  代表第二范数。这意味着, 包含更大数量的递增移动的变换是有利的, 如与对角线移动相关的变换。使用特定 (ad hoc) 规则的另一实施例的示例可以包括: 通过把更大权重给予意欲将路径向中心对角线 (贯穿 K 维空间的从  $(1, 1, \dots, 1)$  到  $(T_1, \dots, T_K)$  的主对角线) 移回的变换而实现的定向偏移。此外, 定向偏移加权可以正比于给定  $\phi(t)$  与主对角线相距的距离, 或者可以仅应用于超出与主对角线相距的特定距离的给定  $\phi(t)$ 。

[0083] 在另一实施例中, MPVA 可以计算  $b_j(\{O_{\phi(t)}\})$ 。即使特征矢量  $\{O_{t_1}^1 O_{t_2}^2, \dots, O_{t_K}^K\}$  可以形成同一类别, 如果给定了它们从同一状态 j 中出现, MPVA 也可以假设它们是独立的, 以便计算从 HMM 中发射的矢量的联合似然。因此, 联合概率可以是

$$[0084] b_j(\{O_{\phi(t)}\}) = [b_j(O_{t_m}^m) \dots b_j(O_{t_n}^n)]^{\frac{1}{r}} \quad (13)$$

[0085] 其中  $(O_{t_m}^m, \dots, O_{t_n}^n)$  是集  $S_{\phi(t)}$  中的所有特征矢量,  $b_j(O_{t_i}^i)$  是 HMM 的状态 j 发射概率 (给定 HMM, 由状态 j 发射的矢量  $O_{t_i}^i$  的概率), 以及 r 是集  $S_{\phi(t)}$  的基数。使用  $1/r$  次幂的几何平均值

对 HMM 状态所发射的  $r$  个矢量的使用进行归一化, 可与单个矢量似然相当。这种归一化考虑了: 由于噪声、时间规整等, 在一时刻并不发射所有特征矢量或观测值。因此, MPVA 可以使用在标准 HMM 中定义的  $a_{ij}$  和  $\pi_i$ 。如果从正确的 HMM 模型  $\lambda$  的实际状态  $j$  中发射  $O_{t_i}^i$ , 则 MPVA 可以期望  $b_j(O_{t_i}^i)$  具有比从不正确的模型的状态  $j$  中发射  $O_{t_i}^i$  的情况高的值。式 (13) 中给出的多模式联合似然可以增强关于正确模型和不正确模型的似然之间的对比度。因此, 在语音识别中存在相比于单独解码提高了准确性的改进。

[0086] MPVA 执行终止步骤 455 以停止递归。递归在诸如  $\Phi(t) = (T_1, T_2, \dots, T_K)$  之类的终点处终止。

$$[0087] P^* = \max_{1 \leq i \leq N} \delta_{\phi(T)}(i) \quad (14)$$

$$[0088] q_{\phi(T)}^* = \arg \max_{1 \leq i \leq N} \delta_{\phi(T)}(i) \quad (15)$$

$$[0089] \Phi(t=T)^* = (T_1, T_2, \dots, T_K) = \Phi(t=T) \quad (16)$$

[0090] 之后, MPVA 执行路径回溯步骤 460。

$$[0091] \phi(t)^* = \Gamma_{\phi(t)^*}(q_{\phi(t+1)^*}^*) \quad (17)$$

$$[0092] q_{\phi(t)^*}^* = \Psi_{\phi(t+1)^*}(q_{\phi(t+1)^*}^*) \quad (18)$$

[0093] 其中  $\Phi(t)$  从  $\Phi(T) = (T_1, \dots, T_K)$  改变到  $\Phi(1) = (1, \dots, 1)$ 。直至完成回溯才知道最优时间路径的值  $T$ 。因而在步骤 465,  $q_{\phi(t)^*}^*$  和  $\Phi(t)^*$  的序列 ( $1 \leq t \leq T$ ) 分别给出了最优解码状态序列  $q^*$  和最优解码时间路径  $\Phi^*$ 。经由简单的查找过程来执行回溯。针对终端的时间索引点  $(T_1, \dots, T_K)$ , 确定最终状态是如由成本函数确定的最高概率, 如式 9 中所示。应当理解, 通过使用式 9 中提出的量的变体或近似物, 和 / 或通过限制考虑哪些组值 (如这里所描述的局部或全局约束), 特定成本函数可以采用许多形式。根据该确定, 获知前驱状态以及相应的前驱时间索引点。然后, 处理器简单地获取在该时间索引点处的该先前状态的数据元素, 并查找下一先前状态和下一先前前驱时间索引点, 依此类推, 直至到达点  $(1, 1, \dots, 1)$ 。

[0094] 参照图 5, 示出了针对  $K = 2$  (词“Voice Dialer”) 及 9 态从左至右的 HMM 的情况的最优 MPVA 路径 500 的示例。示出了两个模式的最优时间校准  $\Phi^*$ 。图 6 是示出了在多模式维特比算法中使用  $K = 2$  个模式的语音的最优状态序列和最优时间路径的三维网格。具体地, 图 6 示出了沿  $z$  轴的最优 HMM 状态序列  $q^*600$ , 而在  $x$  和  $y$  轴上示出了最优时间校准  $\Phi^*$ 。

[0095] 此外, MPVA 可以使用全局路径约束 (GPC) 来显著降低实现 MPVA 的计算复杂度。可以预料, 最优路径将会靠近于对角线, 并且并不是所有时间索引点都需要填充有状态成本度量, 因而可以使用全局路径约束。如果通常以等同比率使用 (或消耗、或处理) 观测值, 则将会遍历对角线路径。将通过重复发音的观测值的相对时间规整产生相对于另一序列的一个序列中的观测值的延迟使用, 并通过观测值序列的不同长度来暗示该延迟使用。可以基于观测值序列的持续时间差来预先确定来自通过  $K$  维时间网格的对角线的最大偏离 (divergence), 但是也可以使用其他度量来确定全局约束。然而, 由于利用了全局约束, 会难以对识别性能进行评估。由于通过式 (13) 的联合解码获得的鲁棒特性, 可以期望性能

(进而识别准确率) 随模式个数 K 的增加而增加。

[0096] 在另一步骤 470, 作为实施 MPVA 的处理器的一部分的解码器可以向系统输出接口提供最优(最大)联合概率和最优状态序列。可以向语音识别系统提供表示最优路径或状态序列的数字数据, 从而使用模拟到数字转换器, 将数字数据变换为模拟电信号。此外, 系统输出接口可以是扬声器或将模拟电信号变换为音频信号的一些其他声换能器设备。系统输出接口可以向呼叫者提供解码语音的音频信号, 以确认呼叫者语音发音的准确率。

[0097] 图 8 是使用多模式维特比算法从多个信号观测值中检测信号的示例方法 800 的流程图。示例方法中的 810 处可以是针对 K 个信号重复中的每一个, 接收多个时间序列信号观测值集。此外, 每个信号观测值集与具有时间索引点的 K 维时间网格的相应维度相关联。该方法中的另一步骤 820 可以是, 在多个时间索引点的每一个处, 针对隐式马尔可夫模型(HMM) 状态集中的每个状态, 利用处理器来计算状态成本度量。此外, 针对每个状态和每个给定的时间索引点, 状态成本度量计算提供了最可能的前驱状态和相应的最可能的前驱时间索引点。此外, 针对给定的时间索引点, 通过计算与每个可能的前驱时间索引点处的每个可能先前状态相关联的成本度量, 并选择每个状态的最低成本度量, 确定每个状态的状态成本度量。此外, 针对给定的可能前驱状态, 状态成本度量仅基于与在从给定前驱时间索引点移到多个时间索引点中的给定时间索引点时递增的维度相关联的观测值。

[0098] 该方法中的其他步骤可以是, 使用所计算的状态成本度量 830 来确定状态序列, 以及确定 HMM 840 的相应累积概率测量。该方法可以以递归方式, 重复计算针对时间索引点的每个状态的状态成本度量, 以确定最可能序列, 直至到达 K 维网格 850 中的终点。该方法可以基于多个 HMM 的相应累积概率测量来标识最可能的 HMM。此外, 确定最可能状态序列包括: 标识终端时间索引点的最终状态的最低状态成本度量。

[0099] 用于该方法的信号观测值是信号特征矢量, 并且可以从由梅尔倒谱系数特征矢量、线性预测编码系数、谱密度、谱能量、噪声比、声长、相对功率、和滤波器匹配构成的组中进行选择。每个观测值组具有相应的持续时间, 预定距离基于相应的持续时间之差。

[0100] 此外, 使用计算状态成本度量来确定状态序列可以包括: 基于状态成本度量的通过时间索引点的回溯。此外, 计算针对给定状态的给定状态成本度量基于: (i) 与所有候选前驱时间索引点相关联的所有状态的状态成本度量; (ii) 从每个候选前驱时间索引点变换到给定状态的变换概率; (iii) 从相应的候选前驱时间索引点变换的相应变换概率; 以及 (iv) 从状态集中的状态发射的观测值的联合概率。

[0101] 此外, 所确定的状态序列确定观测值集的对齐。关于与通过 K 维空间的对角线相距的预定距离来确定多个时间索引点。

[0102] 基于确定 MPVA 相比于其他方法和 / 或算法的性能, 针对说话者无关的孤立词识别(IWR) 实验, 进行包括基本维特比算法(VA) 的实验 A1、A2、A3, 将其结果呈现于本公开中。进行实验 A1, 从而 A1 使用 K 个模式, 选择 K 个模式的最佳(最大)似然。给定  $O_{i:T_i}^1, O_{i:T_2}^2, \dots, O_{i:T_K}^K$  作为属于同一类别的单独模式, 获得联合似然评分为

$$[0103] \theta_j = \max_{1 \leq i \leq K} P(O_{i:T_i}^j; \lambda_j),$$

[0104] 其中  $\lambda_j$  是干净字模型, VA 用于计算  $P(O_{i:T_i}^j; \lambda_j)$ 。选择模式  $j^* = \operatorname{argmax}_j \theta_j$ 。此外, 选择具有最高似然的模式。实验 A2 是约束多模式维特比算法(CMPVA) 的版本(用于

实验 A2 的特定版本被称为 CMPVA-1), 以及实验 A3 使用 MPVA。使用式 (11) 计算  $P(\phi(t)/\phi(t-1))$ , 以及使用式 (13) 计算  $b_j(\{o_{\phi(t)}\})$ 。此外, 使用具有全局路径约束的 MPVA 来进行实验 A3(GPC)。

[0105] 使用 IISc-BPL 数据库来进行实验, 该 IISc-BPL 数据库包括由 36 位成年女性和 34 位成年男性说话者所说的 75 词词表, 每个词由同一说话者重复三次, 以 8kHz 采样率进行数字化。IISc-BPL 数据库是用于语音拨号器应用的印度口音的英语数据库。该数据库包括在使用 5 个不同记录信道 (PSTN 电话 (8kHz 采样)、无绳局域电话 (16kHz 采样)、直接麦克风 (16kHz 采样)、爱立信 (GSM) 移动电话 (8kHz 采样)、混响室电话 (索尼) (8kHz 采样)) 的实验环境中记录的 36 位成年女性和 34 位成年男性说话者所说的英语孤立词, 英语 TIMIT 句、自然语句 (针对不同说话者而不同)。

[0106] 词表包括用于语音拨号器应用中的大量的发音易混词。使用部分 K 平均 (SKM) 算法, 针对干净语音来训练从左到右的 HMM。25 位男性和 25 位女性说话者用于训练, 每位说话者对于每个词重复三次。在干净和嘈杂的情况下, 针对 20 位未见过的说话者 (11 位女性和 9 位男性) 来测试 MPVA。测试词是以每个信噪比 (SNR) 针对每个说话者的每个词的三种模式。针对突发噪声影响的语音来运行实验。对于所有三种模式, 将突发噪声以 -5dB、0dB、5dB SNR( 局部 ) 添加到每个词的 10% 的帧。( 剩余帧是干净的 ; -5dB 到 +5dB 的范围指示对于受噪声影响的帧的轻微到严重降级。) 突发噪声随机地出现在具有均匀概率分布的所说的词中的任何地方。使用 MFCC、 $\Delta$ MFCC 和  $\Delta$ 2MFCC, 而无需其能量分量 (36 维)。忽略能量分量, 进行倒谱均值减法。针对每个词模型, 使用可变数量的状态 ; 即, 正比于训练模型的平均持续时间, 针对话音的每一秒, 分配 8 个 HMM 状态, 每个状态 3 个高斯混合。还针对如机枪噪声和串音噪声之类的其他种类的瞬时噪声 ( 以各种 SNR 取自 NOISEX 92 数据库 ) 来进行实验。

[0107] 表 1 中总结了实验结果, 比较了针对  $K = 2$  个模式, 在实验 VA、A1、A2 和 A3 中对于干净和嘈杂的语音的 ASR 百分比准确率。使用式 (13) 来计算  $b_j(\{o_{\phi(t)}\})$ 。使用式 (11) 来计算  $P(\phi(t)/\phi(t-1))$ 。将突发噪声以变化的 SNR 添加到 10% 的帧。将串音噪声和机枪噪声添加到整个话音模式。表 1 的列标题如下 :VA 维特比算法 ; 实验 A1, 使用 VA 的两个模式中的最佳模式 ; 实验 A2, MPDTW+CMPVA-1 ; 实验 A3, MPVA ; 实验 A3(GPC) , 具有全局路径约束的 MPVA。

[0108] 使用具有  $K = 2$  个测试模式的 MPVA 的实验 A3 的结果给出了相比于 VA 有 2.3% 的 ASR 准确率的提高。有趣的是, 实验 A1 提供了相比于 VA 基准, 分别针对干净和嘈杂的语音 ( 以 -5dB SNR, 10% 的突发噪声 ) 有  $\approx 0.2\%$  和 3% 的略微提高。这表明, 使用多模式确实有益, 只是, 似然最大化能力较弱。实验 A3 也明显好于只考虑两个模式中较好的模式用于识别的实验 A1。然而, 针对干净的语音, 在实验 A2 和 A3 中的 ASR 准确率并没有太大差异。

[0109] 现在考虑具有突发噪声的语音的情况。表 1 示出了以 -5dB SNR、10% 的突发噪声语音, 使用所提出的 MPVA 的词出错率比使用单个模式 VA 时的词出错率降低了 28.5%。此外, 表 1 示出了, 与使用先前提出的 CMPVA-1 ( 实验 A2 ) 相比, 使用 MPVA ( 实验 A3 ) 的词出错率有 19.9% 的大幅降低。使用针对统计显著性差异的 McNemar 测试找出, 实验 A2 和 A3 对于突发噪声情况的实验结果差异在统计学上非常显著, 实验 VA 和 A3 亦是如此。

[0110] 类似地, 表 1 示出了针对以 5dB 或 10dB SNR 的诸如串音噪声和机枪噪声之类的其

他嘈杂语音,与实验 VA 相比,实验 MPVA 在 ASR 准确率上有提高。然而,当使用串音和机枪噪声时,实验 A2 和 A3 的性能没有显著差异。

[0111] 当以 5dB SNR 的加性高斯白噪声 (AWGN) 影响 100% 的所有测试语音模式时,进行实验的变体。VA 给出了 22.89% 的百分比准确率。实验 A3 给出了接近于 VA 准确率但稍差些的 21.73% 的准确率。然而,当以 -5dB SNR 添加 AWGN 时,VA 给出了 3.73% 的准确率,实验 3 的准确率提高到 4.09%。因此,对于像 AWGN 的稳定噪声,MPVA 提供了与 VA 相类似性能。这优于在 AWGN[5] 影响语音时给出比 VA 差得多的性能的 CMPVA-1。

[0112] 从表 1 中还可以看出,GPC(实验 A3(GPC)) 的使用并没有显著改变 MPVA 的性能。由于 GPC 的使用将计算复杂度降低了多于 50%,因而实验证明有理由使用 GPC。

[0113] 表 1

	VA	A1	A2	A3	A3 (GPC)
[0114]	干净	89.70	89.87	91.78	92.00
	-5 dB, 10% 突发噪声	57.13	60.33	61.73	69.33
	0 dB, 10% 突发噪声	61.49	64.29	66.18	72.42
	5 dB, 10% 突发噪声	67.38	69.49	72.47	75.93
	5 dB 串音噪声	44.27	44.33	49.84	49.07
	10 dB 串音噪声	59.73	60.71	65.51	64.93
	5 dB 机枪噪声	66.71	68.09	74.56	74.28
	10 dB 机枪噪声	71.47	72.44	78.47	77.82
					77.96

[0115] 考虑用于实验 A1、A2 和 A3 的不同算法的计算复杂度。维特比算法的复杂度可以包括计算  $b_j(0t)$  (这是具有指数项的高斯混合),其作为 1 个 Es 单元,以及 HMM 具有 N 个状态。定义被称为 MAC(乘加)的单元。一个 MAC 是一个加法和一个乘法运算的组合。为了简单,假设一个最大 (max) 运算等于一个加法运算。在格式结构中的每个节点处,存在 {N-1 个 max 运算, N+1 个乘法, 和 1 个 Es 单元}  $\approx$  {N 个 MAC, 1 个 Es}。维特比算法的总复杂度是 {T. N2 个 MAC, T. N 个 Es},其中 T 是模式中的帧数。

[0116] MPVA 的复杂度可以包括计算联合概率  $b_j(\{0_{\phi(t)}\})$ ,作为 X 个 Es 单元。式 (11) 的  $P(\phi(t)/\phi(t-1))$  具有一个乘法运算。在网式结构 (式 (9)) 中的每个节点处,存在 {N-1 个 max 运算, 2N+1 个乘法, X 个 Es}。为了简单,假设 T1 = T2 = ... = TK = T。MPVA 的总复杂度为 {TK. N2 个 MAC, TK. N2 个乘法, TK. N. X 个 Es}。

[0117] 这些实施例描述了用于使用多模式维特比算法 (MPVA)、针对 K 个模式集来联合确定单个最优状态序列的方法。该方法包括:在确定穿过 (K+1) 维网格的最优时间路径的过程中,选择局部连续性约束 (LCC) (如果有的话)。基于 MPVA, ASR 准确率显著高于单个模式 VA 的情况。在有噪声的情况下,MPVA 技术也优于 CMPVA 技术。MPVA 程式具有可应用于许多需要基于 HMM 的模式匹配的鲁棒性的其他问题的通用性。与 VA 扩展到联合地识别多个模式相类似,可以扩展前向和后向算法,以联合地识别多个模式。

[0118] 图 7 是示出了被设置用于使用 MPVA 的模式识别系统的示例计算设备 700 的框图。在非常基本的配置 701 中,计算设备 700 典型地包括一个或多个处理器 710 和系统存储器

720。存储器总线 730 可以用于在处理器 710 与系统存储器 720 之间通信。根据所期望的配置,处理器 710 可以是任意类型的,包括但不限于微处理器 ( $\mu$  P)、微控制器 ( $\mu$  C)、数字信号处理器 (DSP) 或其任意组合。处理器 710 可以包括一级或多级高速缓存 (例如,一级高速缓存器 711 和二级高速缓存器 712)、处理器核 713、以及寄存器 714。处理器核 713 可以包括算术逻辑单元 (ALU)、浮点单元 (FPU)、数字信号处理核心 (DSP 核) 或其任意组合。存储控制器 715 也可以与处理器 710 一起使用,或者在一些实施方式中,存储控制器 715 可以是处理器 710 的内部部件。

[0119] 根据所期望的配置,系统存储器 720 可以是任意类型的,包括但不限于易失性存储器 (如 RAM)、非易失性存储器 (如 ROM, 闪存等) 或其任意组合。系统存储器 720 典型地包括操作系统 721、一个或多个应用程序 722 和程序数据 724。根据本公开,应用程序 722 包括控制输入处理算法 723, 被设置为向电子电路提供输入。程序数据 724 包括控制输入数据 725, 如将在以下进一步描述的, 该控制输入数据 725 用于将电路的功耗最小化。在一些示例实施例中, 应用程序 722 可以被设置为在操作系统 721 上以程序数据 724 操作, 使得将电子电路的功耗最小化。这里所描述的基本配置在图 7 中由虚线 701 内的部件来图示。

[0120] 计算设备 700 可以具有额外特征或功能以及额外接口, 以有助于基本配置 701 与任意所需设备和接口之间进行通信。例如, 总线 / 接口控制器 740 可以有助于基本配置 701 与一个或多个数据存储设备 750 之间经由存储接口总线 741 进行通信。数据存储设备 750 可以是可拆除存储设备 751、不可拆除存储设备 752 或其组合。可拆除存储和不可拆除存储设备的示例包括磁盘设备 (如软盘驱动器和硬盘驱动器 (HDD))、光盘驱动器 (如紧凑型光盘 (CD) 驱动器或数字通用盘 (DVD) 驱动器)、固态驱动器 (SSD) 以及磁带驱动器, 这仅仅是极多例子中的一小部分。示例计算机存储介质可以包括以任意信息存储方法和技术实现的易失性和非易失性、可拆除和不可拆除介质, 如计算机可读指令、数据结构、程序模块或其他数据。

[0121] 系统存储器 720、可拆除存储器 751 和不可拆除存储器 752 均是计算存储介质的所有示例。计算存储介质包括但不限于 RAM、ROM、EEPROM、闪存或其他存储器技术, CD-ROM、数字通用盘 (DVD) 或其他光存储器, 磁卡带、磁带、磁盘存储器或其他磁存储设备, 或可以用于存储所需信息并可以由计算设备 700 访问的任意其他介质。任何这种计算机存储介质可以是设备 700 的一部分。

[0122] 计算设备 700 还可以包括接口总线 742, 以有助于各种接口设备 (例如, 输出接口、外设接口和通信接口) 经由总线 / 接口控制器 740 与基本配置 701 进行通信。示例输出接口 760 包括图形处理单元 761 和音频处理单元 762, 其可被配置为经由一个或多个 A/V 端口 763 与多种外部设备 (如显示器或扬声器) 进行通信。示例外围设备接口 770 包括串行接口控制器 771 或并行接口控制器 772, 它们可被配置为经由一个或多个 I/O 端口 773 与外部设备 (如输入设备 (例如, 键盘、鼠标、笔、语音输入设备、触摸输入设备等)) 或其他外设设备 (例如, 打印机、扫描仪等) 进行通信。示例通信接口 780 包括网络控制器 781, 其可以被配置为经由一个或多个通信端口 782 与一个或多个其他计算设备 790 通过网络通信连接进行通信。通信连接可以是通信介质的一个示例。通信介质典型地可以由调制数据信号 (如载波或其他传输机制) 中的计算机可读指令、数据结构、程序模块或其他数据来体现, 并可以包括任意信息传送介质。“调制数据信号”可以是通过设置或改变一个或多个特性而在该

信号中实现信息编码的信号。例如,但并非限制性地,通信介质可以包括有线介质(如有线网络或直接布线连接)、以及无线介质(例如声、射频(RF)、微波、红外(IR)和其他无线介质)。这里所使用的术语计算机可读介质可以包括存储介质和通信介质。

[0123] 计算设备700可以实现为小体积便携式(或移动)电子设备的一部分,如蜂窝电话、个人数字助理(PDA)、个人媒体播放器设备、无线web浏览设备、个人手机设备、专用设备或可以包括任意上述功能的混合设备。计算设备700也可以实现为个人计算机,包括膝上型计算机和非膝上型计算机配置。

[0124] 通常,应当理解,这里所描述的电路可以通过使用集成电路开发技术的硬件、或通过一些其他方法、或可以在软件环境中被排序、特征化并连接以实现这里所描述的不同功能的硬件和软件对象的组合来实现。例如,本申请可以使用通过易失性或非易失性存储器运行软件应用的通用或专用处理器来实现。此外,硬件对象可以使用电信号进行通信,其中信号的状态表示不同数据。

[0125] 还应理解,这里所描述的这个和其他设置的目的仅为示意性。因而,本领域技术人员将会理解,可以使用其他设置和其他元件(例如,及其、接口、功能、命令和功能组等)来替代,根据期望结果,可以省略一些元件。此外,这里所描述的许多元件是可以实现为离散或分布式组件或以任何适合方式和位置与其他组件相结合的功能实体。

[0126] 本公开并不限于该申请中描述的特定实施例,这些特定实施例意在说明各个方面。可以在不偏离本公开的精神和范围的情况下做出许多修改和变化,这将对本领域技术人员来说显而易见。通过上述描述,除了这里所列举的之外,在本公开的范围内的功能上等同的方法和设备对于本领域技术人员来说也是显而易见的。这些修改和变化意在落入所附权利要求的范围之内。本公开仅通过所附权利要求、以及这些权利要求等同物的全部范围进行限定。要理解,本公开并不限于特定的方法、试剂、化合物成分或生物学系统,当然,它们可以变化。还要理解,这里所使用的术语仅出于描述特定实施例的目的,并不意在做出限定。

[0127] 至于本文中任何关于多数和/或单数术语的使用,本领域技术人员可以从多数形式转换为单数形式,和/或从单数形式转换为多数形式,以适合具体环境和应用。为清楚起见,在此明确声明单数形式/多数形式可互换。

[0128] 本领域技术人员应当理解,一般而言,所使用的术语,特别是所附权利要求中(例如,在所附权利要求的主体部分中)使用的术语,一般地应理解为“开放”术语(例如,术语“包括”应解释为“包括但不限于”,术语“具有”应解释为“至少具有”等)。本领域技术人员还应理解,如果意在所引入的权利要求中标明具体数目,则这种意图将在该权利要求中明确指出,而在没有这种明确标明的情况下,则不存在这种意图。例如,为帮助理解,所附权利要求可能使用了引导短语“至少一个”和“一个或多个”来引入权利要求中的特征。然而,这种短语的使用不应被解释为暗示着由不定冠词“一”或“一个”引入的权利要求特征将包含该特征的任意特定权利要求限制为仅包含一个该特征的实施例,即便是该权利要求既包括引导短语“一个或多个”或“至少一个”又包括不定冠词如“一”或“一个”(例如,“一”和/或“一个”应当被解释为意指“至少一个”或“一个或多个”);在使用定冠词来引入权利要求中的特征时,同样如此。另外,即使明确指出了所引入权利要求特征的具体数目,本领域技术人员应认识到,这种列举应解释为意指至少是所列数目(例如,不存在其他修饰语的

短语“两个特征”意指至少两个该特征，或者两个或更多该特征）。另外，在使用类似于“A、B 和 C 等中至少一个”这样的表述的情况下，一般来说应该按照本领域技术人员通常理解该表述的含义来予以解释（例如，“具有 A、B 和 C 中至少一个的系统”应包括但不限于单独具有 A、单独具有 B、单独具有 C、具有 A 和 B、具有 A 和 C、具有 B 和 C、和 / 或具有 A、B、C 的系统等）。在使用类似于“A、B 或 C 等中至少一个”这样的表述的情况下，一般来说应该按照本领域技术人员通常理解该表述的含义来予以解释（例如，“具有 A、B 或 C 中至少一个的系统”应包括但不限于单独具有 A、单独具有 B、单独具有 C、具有 A 和 B、具有 A 和 C、具有 B 和 C、和 / 或具有 A、B、C 的系统等）。本领域技术人员还应理解，实质上任意表示两个或更多可选项目的转折连词和 / 或短语，无论是在说明书、权利要求书还是附图中，都应被理解为给出了包括这些项目之一、这些项目任一方、或两个项目的可能性。例如，短语“A 或 B”应当被理解为包括“A”或“B”、或“A 和 B”的可能性。

[0129] 另外，在以马库什组描述本公开的特征或方案的情况下，本领域技术人员应认识到，本公开由此也是以该马库什组中的任意单独成员或成员子组来描述的。

[0130] 本领域技术人员应当理解，出于任意和所有目的，例如为了提供书面说明，这里公开的所有范围也包含任意及全部可能的子范围及其子范围的组合。任意列出的范围可以被容易地看作充分描述且实现了将该范围至少进行二等分、三等分、四等分、五等分、十等分等。作为非限制性示例，在此所讨论的每一范围可以容易地分成下三分之一、中三分之一和上三分之一等。本领域技术人员应当理解，所有诸如“直至”、“至少”、“大于”、“小于”之类语言包括所列数字，并且指代了随后可以如上所述被分成子范围的范围。最后，本领域技术人员应当理解，范围包括每一单独数字。因此，例如具有 1 ~ 3 个单元的组是指具有 1、2 或 3 个单元的组。类似地，具有 1 ~ 5 个单元的组是指具有 1、2、3、4 或 5 个单元的组，以此类推。

[0131] 尽管已经在此公开了多个方案和实施例，但是本领域技术人员应当明白其他方案和实施例。这里所公开的多个方案和实施例是出于说明性的目的，而不是限制性的，本公开的真实范围和精神由所附权利要求表征。

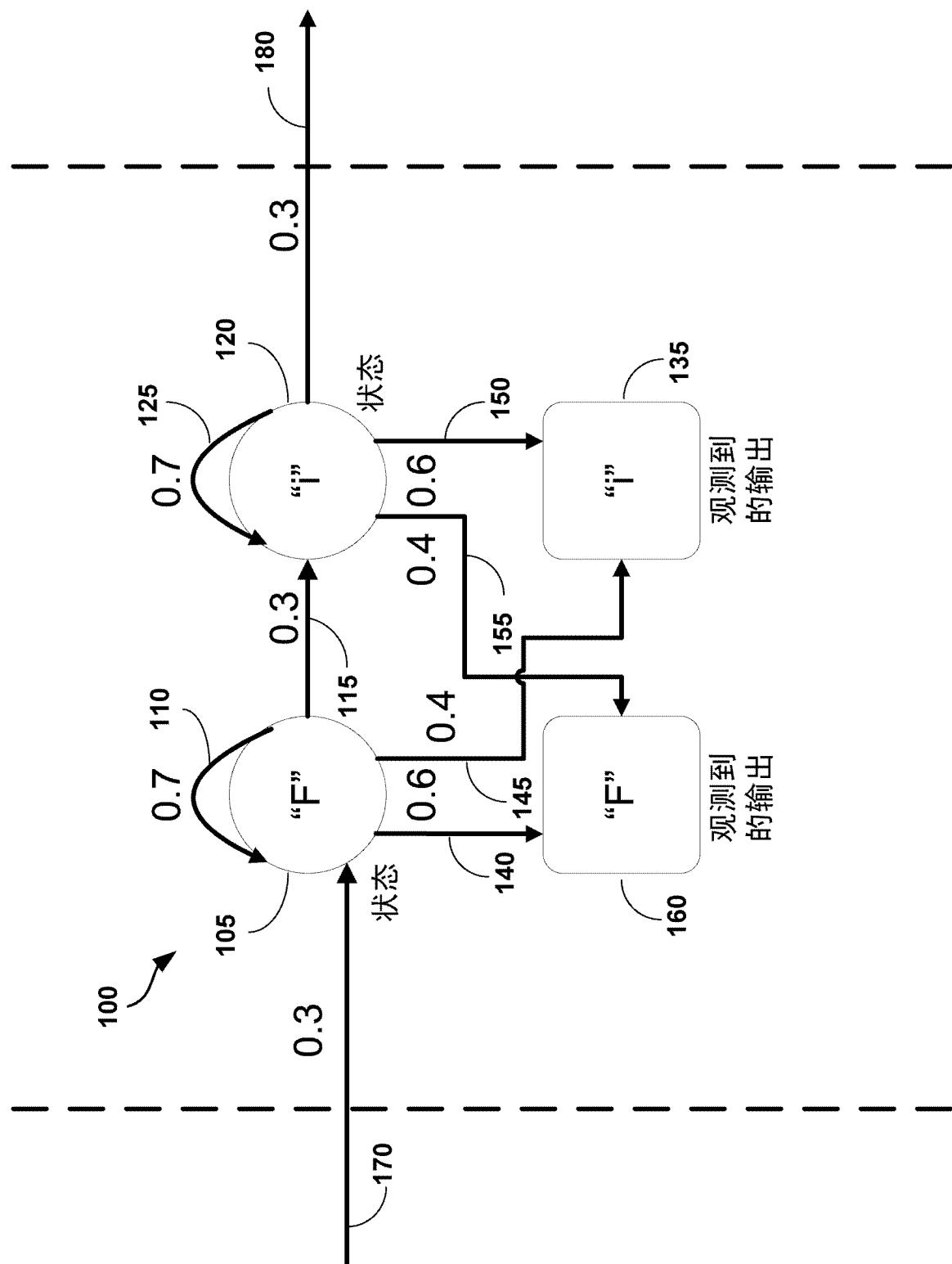


图 1

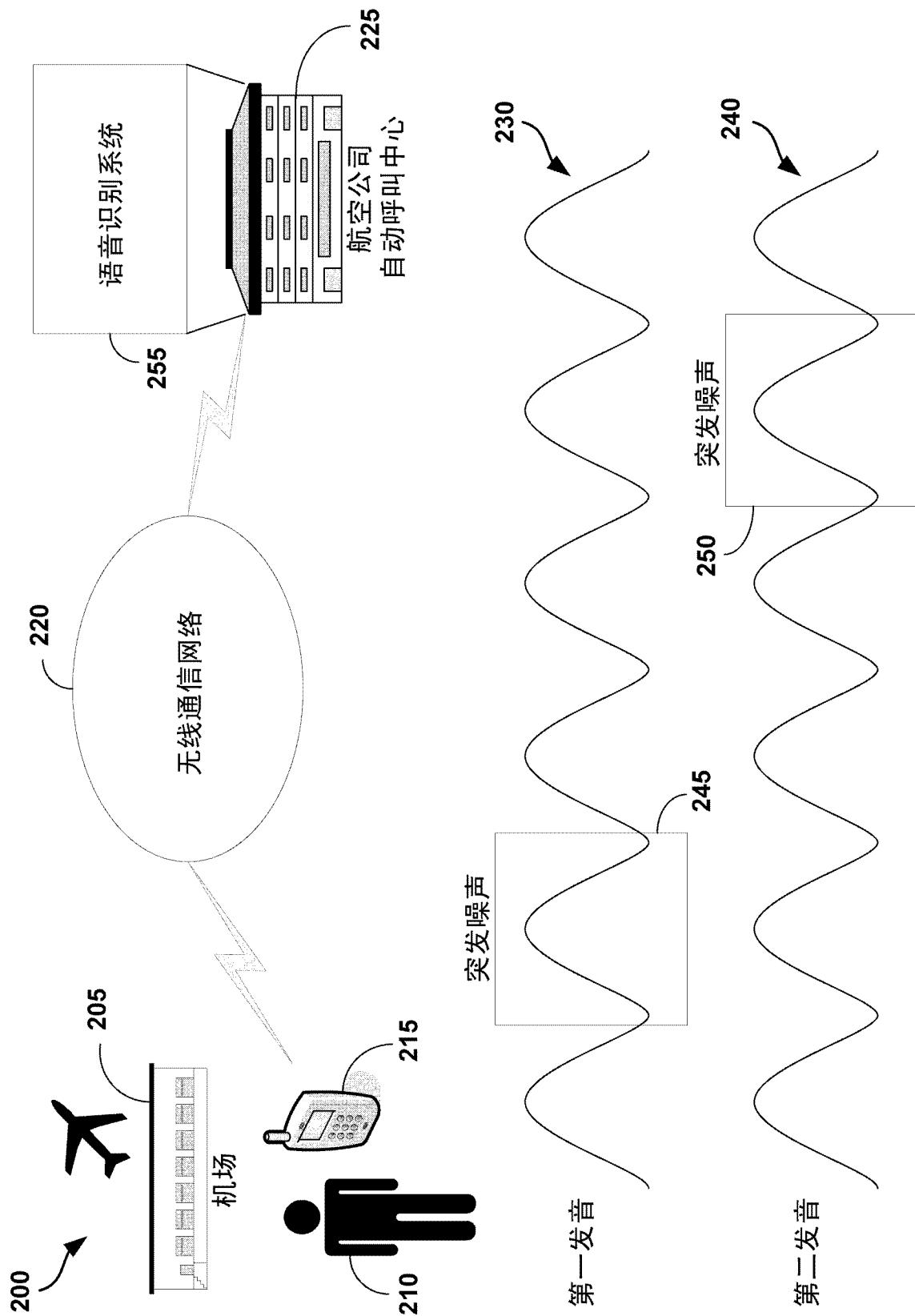


图 2

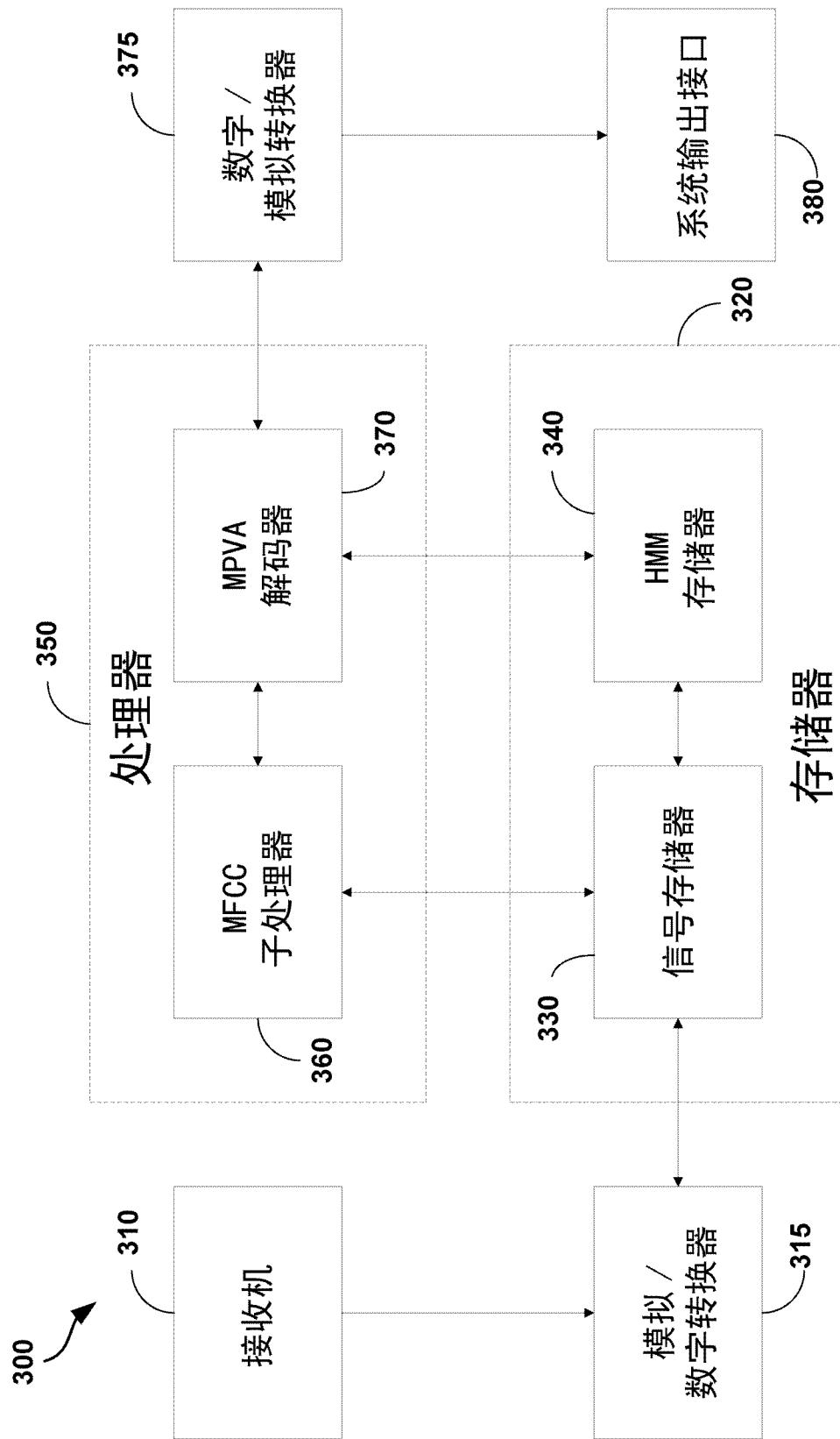


图 3

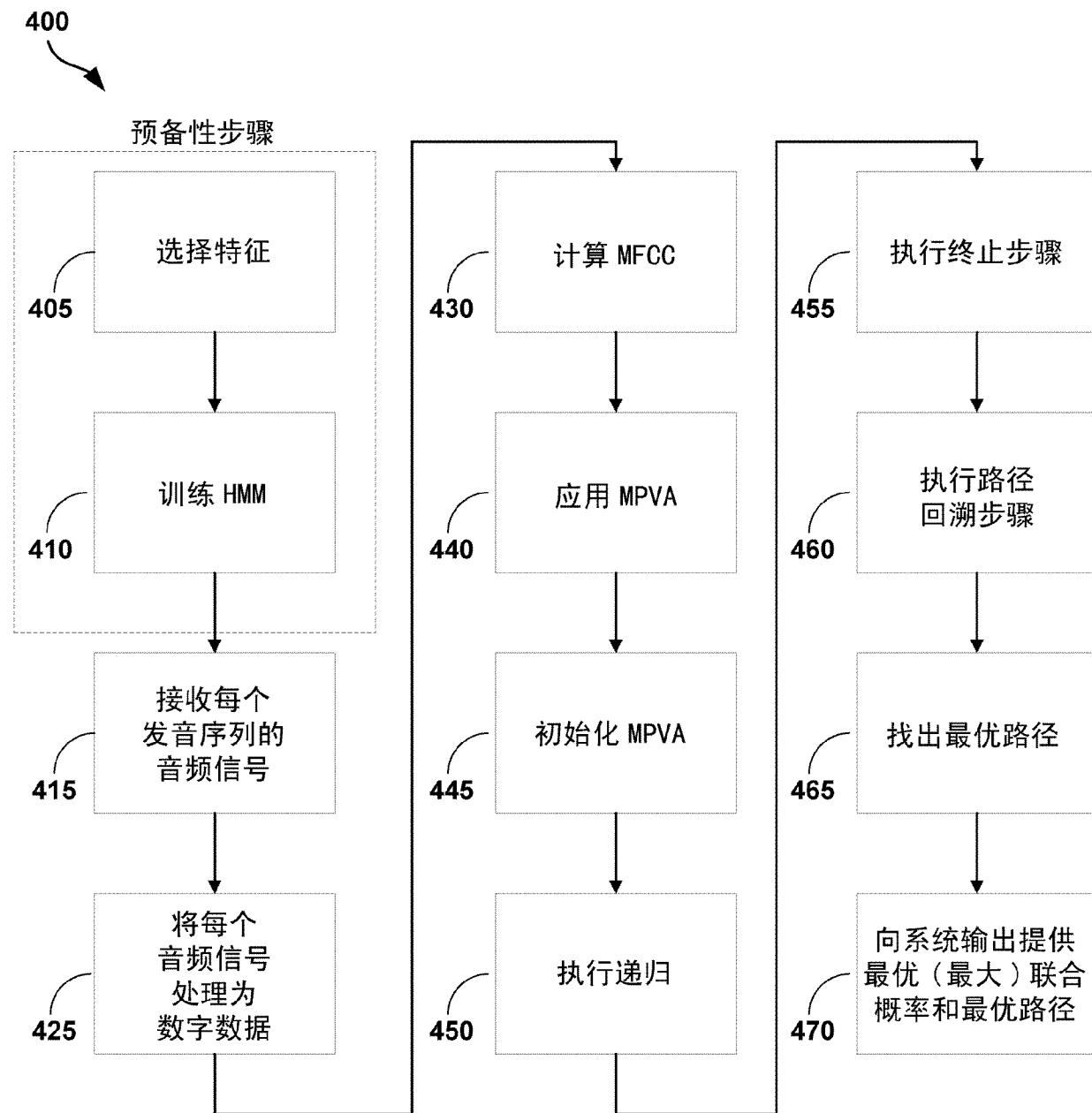


图 4

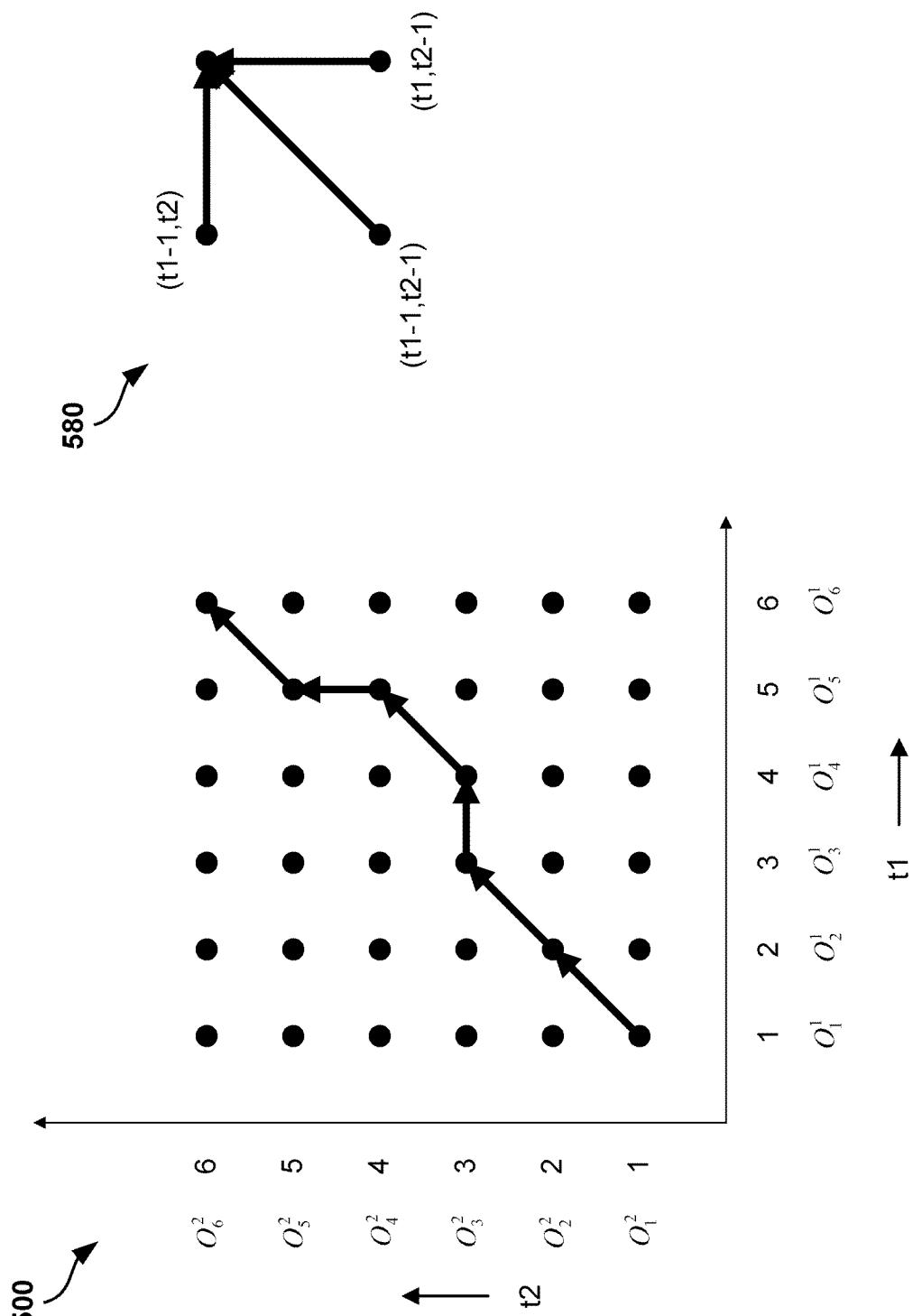
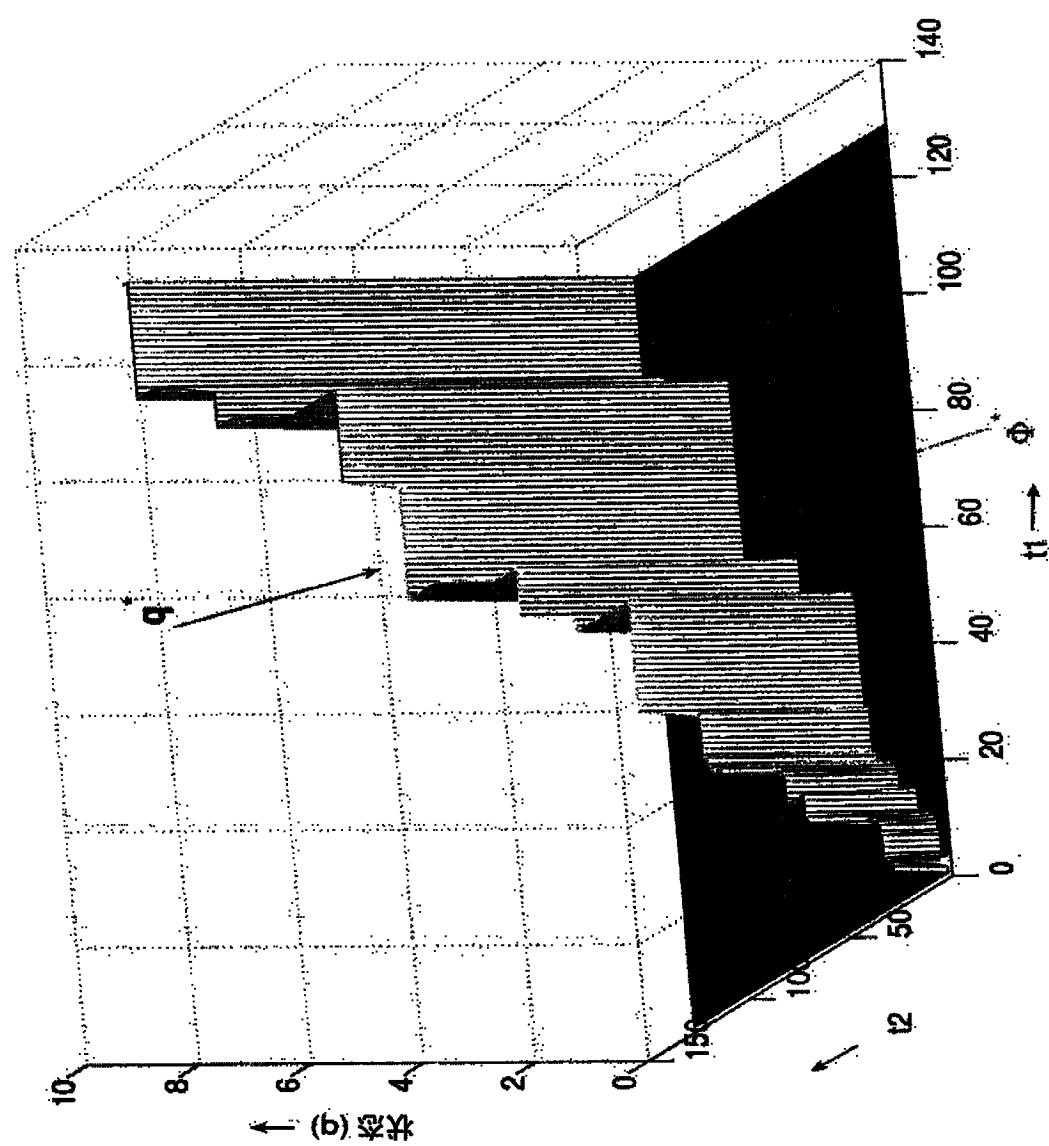


图 5



600  
↗

图 6

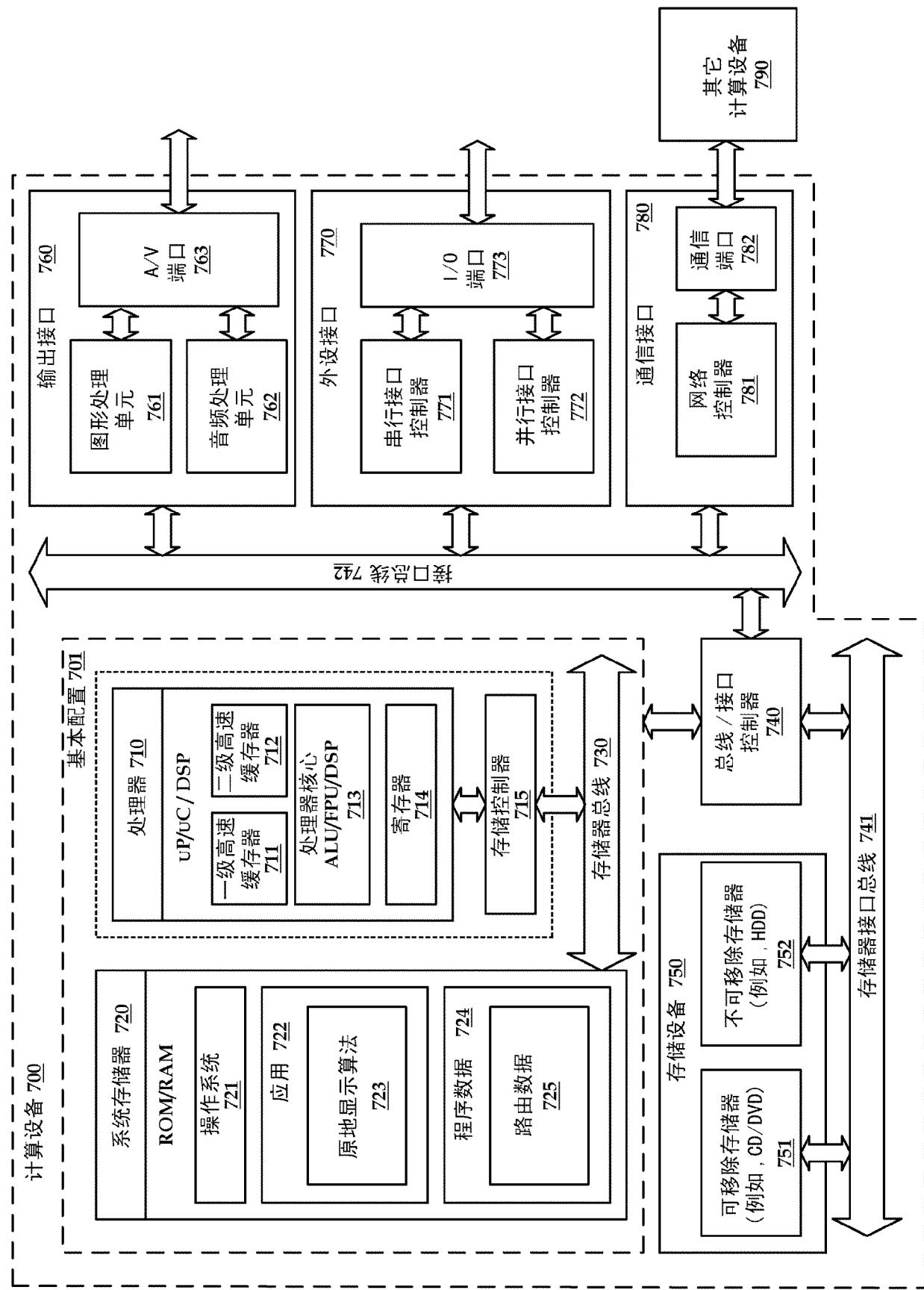


图 7

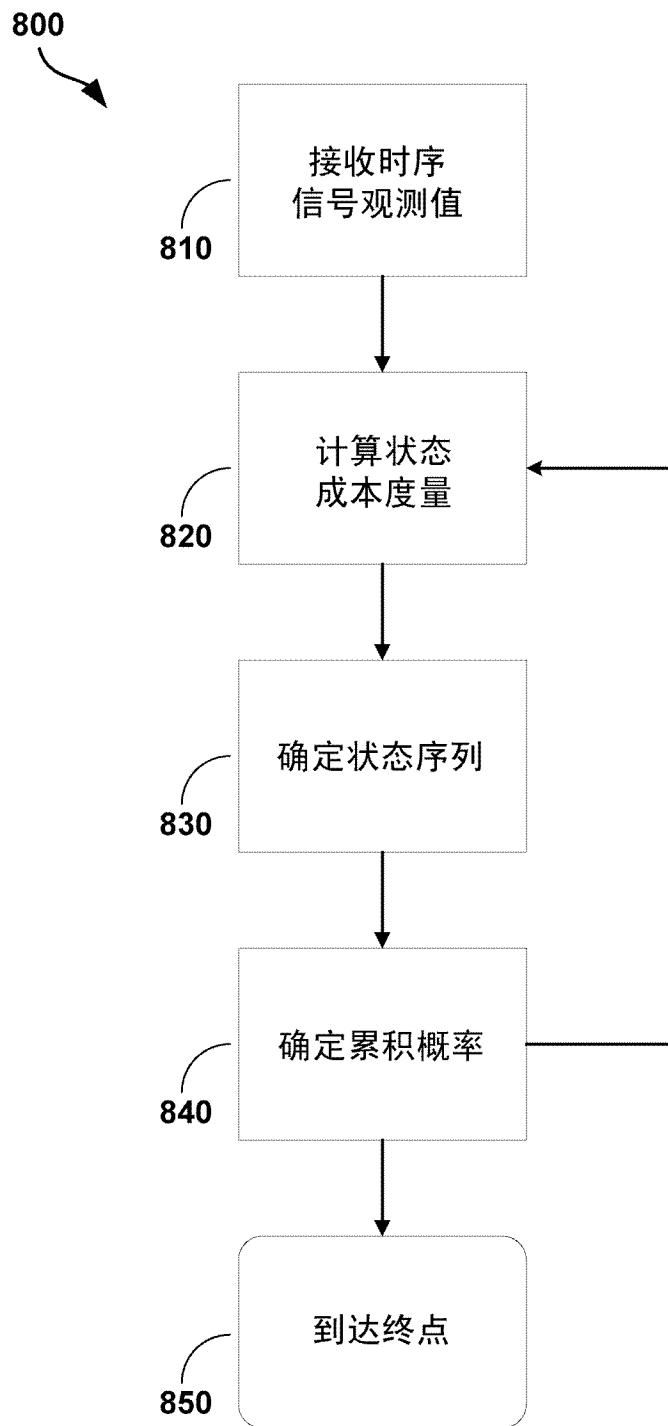


图 8