



[12] 发明专利说明书

[21] ZL 专利号 97193175.5

[45] 授权公告日 2003 年 8 月 27 日

[11] 授权公告号 CN 1119795C

[22] 申请日 1997.12.22 [21] 申请号 97193175.5

[30] 优先权

[32] 1996.12.26 [33] JP [31] 347312/1996

[86] 国际申请 PCT/JP97/04755 1997.12.22

[87] 国际公布 WO98/29864 日 1998.7.9

[85] 进入国家阶段日期 1998.9.18

[71] 专利权人 索尼公司

地址 日本东京都

[72] 发明人 近藤哲二郎

审查员 杨 叁

[74] 专利代理机构 北京市柳沈律师事务所

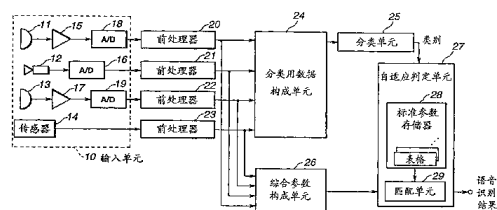
代理人 马 莹

权利要求书 2 页 说明书 14 页 附图 9 页

[54] 发明名称 语音识别装置和识别方法以及语音学习装置和学习方法

[57] 摘要

从输入单元(10)输出多种数据,例如,用户发出的语音数据、通过对用户口部摄象而获得的图像数据、周围噪声的数据等。这些数据由前处理单元(20-23)分析,从而获得特征参数。在分类用数据构成单元(24)中,根据特征参数构成分类用数据,分类单元(25)将这些数据进行分类。在综合参数构成单元(26)中,根据前处理单元(20-23)输出的特征参数来构成综合参数。在自适应判定单元(27)中,选出与分类单元(25)分类结果相对应的表格。根据登记在所选表格中的标准参数和根据综合参数构成单元(26)输出的综合参数来识别用户发出的语音。因此本发明可提高语音的识别率。



1. 一种从不同类型的输入数据中识别特定对象的语音识别装置,包括:
第一分类设备,用于根据所述不同类型的输入数据的性质将所述不同
5 类型的输入数据分类,每一性质都独立地从其他类型的输入数据中取出;
综合参数构成设备,用于通过综合所述不同类型的输入数据而构成综合参数;
标准参数存储设备,用于存储表格,每个表格携带着指定给由所述第一分类设备确定的一个类别的标准参数;以及
10 识别设备,用于利用所述综合参数和指定给由所述第一分类设备确定的类别的标准参数来识别所述特定对象。
2. 如权利要求1所述的语音识别装置,还包括:
分析设备,用于分析所述不同类型的输入数据,以确定特征参数;以及
15 所述第一分类设备根据所述特征参数执行所述分类。
3. 如权利要求1所述的语音识别装置,还包括:
分析设备,用于根据由所述第一分类设备确定的类别,分析所述不同类型的输入数据,以确定特征参数;
第二分类设备,用于利用每个所述不同类型的输入数据的所述特征参
20 数来执行所述分类;
所述综合参数构成设备通过综合所述不同类型的输入数据的多个特征参数,以确定所述综合参数;
所述标准参数存储设备适合于存储表,每个表指定给由第一分类设备确定的一个类别,并且携带着指定给由所述第二分类设备确定的类别的标准参数;以及
25 所述识别设备适合于利用所述综合参数和列在指定给由所述第一分类设备确定的类别和由所述第二分类设备确定的类别的表中的标准参数,来识别所述给定主体。
4. 如权利要求1所述的语音识别装置,其特征在于:
30 所述特定对象是感兴趣的说话人发出的语音,所述不同类型的输入数据至少包含该语音数据和对说话人嘴部摄象而得到的图像数据。

5. 一种从不同类型的输入数据中识别特定对象的语音识别方法，包括以下步骤：

根据所述不同类型的输入数据的性质，将所述不同类型的输入数据分类，并通过综合所述不同类型的输入数据构成综合参数；以及

5 利用所述综合参数和携带有指定给由所述分类结果确定的一个类别的标准参数，来识别所述特定对象。

6. 一种用于确定标准参数的语音学习装置，所述标准参数被用于从不同类型的输入数据中识别特定对象，所述语音学习装置包括：

10 第一分类设备，用于根据所述不同类型的输入数据的性质，将所述不同类型的输入数据分类；

综合参数构成设备，用于通过综合对所述不同类型的输入数据而构成综合参数；以及

分类设备，用于根据由所述第一分类设备确定的类别，分类每个综合参数。

15 7. 如权利要求 6 所述的语音学习装置，还包括：

分析设备，用于分析所述不同类型的输入数据以确定特征参数；以及所述第一分类设备适合于根据所述特征参数执行所述分类。

8. 如权利要求 6 所述的语音学习装置，还包括：

20 分析设备，用于利用由所述第一分类设备确定的类别，分析所述不同类型的输入数据，以确定特征参数；

第二分类设备，用于利用每个所述不同类型的输入数据的所述特征参数来执行所述分类；

所述综合参数构成设备综合所述不同类型的输入数据的多个所述特征参数，以确定所述综合参数；以及

25 所述分类设备根据由所述第一分类设备确定的类别和由所述第二分类设备确定的类别分类所述综合参数。

9. 一种确定标准参数的语音学习方法，所述标准参数用于从不同类型的输入数据中识别特定对象，所述语音学习方法包括步骤：

30 根据所述不同类型的输入数据的性质，将所述不同类型的输入数据分类，并通过综合所述不同类型的输入数据构成综合参数；以及

按照由所述分类结果确定的类别来分类各个所述综合参数。

语音识别装置和识别方法以及 语音学习装置和学习方法

5

技术领域

本发明涉及识别装置和识别方法以及学习装置和学习方法，具体涉及对于例如声音和物体等不但用其声音及图像数据、还用其它数据进行识别、因而能提高识别精度的识别装置和识别方法以及学习装置和学习方法。

10

背景技术

在已有的例如识别语音的语音识别装置中，是对从送话器输出的语音数据进行(音响)分析及利用该分析结果来识别用户发出的语音的。

但是，在已有的语音识别装置中，由于只用来自送话器的语音数据的分析结果进行语音识别，其识别性能在某种程度上受到限制。

即，在识别说话人的语音时，不但来自送话器的语音数据，而且该说话人的表情和嘴部动作等也是重要因素，因而必须考虑此种等种种因素。

而且，语音识别装置几乎不在如隔音室之类的只把作为识别对象的语音从送话器输入的环境下使用，而往往在存在除此之外的种种噪声的环境下使用。最近，设有语音识别装置的例如导航系统正在实用化，在此场合下，在语音识别装置中，除了应识别的说话人的语音之外，还输入了装备在汽车中的 CD(光盘)播放器的输出声音、引擎的声音、空调器的声音等其它噪声。而从混合有此类噪声的语言数据中只除去噪声是十分困难的。因此，为了提高识别性能，有必要采纳作为识别对象的语音以外的声音的噪声。

25

此外，在已往的语音识别装置中，是对从送话器输出的声音数据进行规定的分析处理而求出其特征参数，然后通过计算在参数空间中该特征参数之间的距离等来识别语音的。但是，要考虑，在识别时，主要的特征参数随语音识别装置的使用环境而变化。

30

发明内容

鉴于以上状况而提出本发明。本发明的目的是提高对语音及其它因素进行识别的识别装置的识别性能。

根据本发明的一方面，提供一种从不同类型的输入数据中识别特定对象的语音识别装置，包括：第一分类设备，用于根据所述不同类型的输入数据的性质将所述不同类型的输入数据分类，每一性质都独立地从其他类型的输入数据中取出；综合参数构成设备，用于通过综合所述不同类型的输入数据而构成综合参数；标准参数存储设备，用于存储表格，每个表格携带着指定给由所述第一分类设备确定的一个类别的标准参数；以及识别设备，用于利用所述综合参数和指定给由所述第一分类设备确定的类别的标准参数来识别所述特定对象。

所述语音识别装置还可以包括：分析设备，用于分析所述不同类型的输入数据，以确定特征参数；以及所述第一分类设备根据所述特征参数执行所述分类。

所述语音识别装置还可以包括：分析设备，用于根据由所述第一分类设备确定的类别，分析所述不同类型的输入数据，以确定特征参数；第二分类设备，用于利用每个所述不同类型的输入数据的所述特征参数来执行所述分类；所述综合参数构成设备通过综合所述不同类型的输入数据的多个特征参数，以确定所述综合参数；所述标准参数存储设备适合于存储表，每个表指定给由第一分类设备确定的一个类别，并且携带着指定给由所述第二分类设备确定的类别的标准参数；以及所述识别设备适合于利用所述综合参数和列在指定给由所述第一分类设备确定的类别和由所述第二分类设备确定的类别的表中的标准参数，来识别所述给定主体。

在所述语音识别装置中，所述特定对象可以是感兴趣的说话人发出的语音，所述不同类型的输入数据至少包含该语音数据和对说话人嘴部摄象而得到的图像数据。

根据本发明的另一方面，提供一种从不同类型的输入数据中识别特定对象的语音识别方法，包括以下步骤：根据所述不同类型的输入数据的性质，将所述不同类型的输入数据分类，并通过综合所述不同类型的输入数据构成综合参数；以及利用所述综合参数和携带有指定给由所述分类结果确定的一个类别的标准参数，来识别所述特定对象。

根据本发明的再一方面，提供一种用于确定标准参数的语音学习装置，

所述标准参数被用于从不同类型的输入数据中识别特定对象，所述语音学习装置包括：第一分类设备，用于根据所述不同类型的输入数据的性质，将所述不同类型的输入数据分类；综合参数构成设备，用于通过综合对所述不同类型的输入数据而构成综合参数；以及分类设备，用于根据由所述
5 第一分类设备确定的类别，分类每个综合参数。

所述语音学习装置还可以包括：分析设备，用于分析所述不同类型的输入数据以确定特征参数；以及所述第一分类设备适合于根据所述特征参数执行所述分类。

所述语音学习装置还可以包括：分析设备，用于利用由所述第一分类
10 设备确定的类别，分析所述不同类型的输入数据，以确定特征参数；第二分类设备，用于利用每个所述不同类型的输入数据的所述特征参数来执行所述分类；所述综合参数构成设备综合所述不同类型的输入数据的多个所述特征参数，以确定所述综合参数；以及所述分类设备根据由所述第一分类设备确定的类别和由所述第二分类设备确定的类别分类所述综合参数。

15 根据本发明的再一方面，提供一种确定标准参数的语音学习方法，所述标准参数用于从不同类型的输入数据中识别特定对象，所述语音学习方法包括步骤：根据所述不同类型的输入数据的性质，将所述不同类型的输入数据分类，并通过综合所述不同类型的输入数据构成综合参数；以及按照由所述分类结果确定的类别来分类各个所述综合参数。

20

附图说明

图 1 是表示应用本发明的导航系统的组成例的方框图。

图 2 是表示应用本发明的语音识别装置第一实施例的组成例的方框图。

图 3 是说明前处理单元 21 的处理用的图。

25 图 4 是表示应用本发明的学习装置第一实施例的组成例的方框图。

图 5 是表示参数空间的图。

图 6 是表示应用本发明的语音识别装置第二实施例的组成例的方框图。

图 7 是表示应用本发明的学习装置第二实施例的组成例的方框图。

图 8 是表示应用本发明的语音识别装置第三实施例的组成例的方框图。

30 图 9 是表示应用本发明的学习装置第三实施例的组成例的方框图。

具体实施方式

图 1 是表示应用本发明的导航系统的实施例的方框图。

该导航系统装备在汽车等中，包括系统控制单元 1、位置测定装置 2、数据库装置 3、输入装置 4 及输出装置 5。该导航系统由例如操作按钮开关
5 或通过对话方式输入的语音来控制。而且，该导航系统也能携带着使用。

系统控制单元 1 在构成该系统的各方框之间进行信息的发送和接收以控制整个系统的动作。位置测定系统 2 接收来自 GPS(全球定位系统)卫星的电波并且用例如陀螺仪或车速传感器的测量设备来测定当前位置。在数据库装置 3 中，记录(存储)着例如多种比例尺的电子化地图信息及其它对导航
10 必要的信息，由此根据来自系统控制单元 1 的指令来检索必要的信息以供给系统控制单元 1。

在输入装置 4 中，装备有对导航系统按规定处理用的按钮开关或操纵杆、输入语音用的送话器、对用户摄象用的 CCD(电荷耦合器件)摄象机、还有测定汽车振动用的加速度传感器、测定湿度和温度用的传感器、以及
15 其它各种传感器等。来自输入装置 4 的与按钮开关或操纵杆操作对应的信号发送到系统控制单元 1。而且，输入装置 4 含有语音识别装置，对输入的声音进行语音识别，然后把该语音识别结果发送到系统控制单元 1。

输出装置 5 含有例如显示图像等用的液晶显示监视器或 CRT(阴极射线管)、输出语音等用的扬声器、由文本信息生成合成语音的语音合成装置等，
20 并且显示从系统控制单元 1 送来的信息之中的例如地图信息和现在位置信息等以及输出语音信息。而且，输出装置 5 在接收到系统控制单元 1 发送来的文本信息时，在语音合成装置中合成并输出与该文本信息对应的语音。

在如上构成的导航系统中，例如，一旦用户说出作为目的地的地名，此语音就在输入装置 4 内装的语音识别装置中进行语音识别，该语音识别
25 结果被供给到系统控制单元 1。一旦系统控制单元 1 接收到作为目的地的地名的语音识别结果，它就根据位置测定装置 2 的输出识别出当前位置，并且参照存储在数据库装置 3 中的地图信息求出从现在位置到目的地的路线。接着，系统控制单元 1 把表示该路线的地图信息供给输出装置 5 进行显示，同时在输出装置 5 内装的语音合成装置中生成并输出引导该路线的合成语
30 音。

因而，在此情况下，用户能容易地到达目的地。

图2表示图1输入装置4内装的语音识别装置第一实施例的组成例子。

输入单元10由送话器11、CCD摄象机12、送话器13、传感器14、放大器15、A/D(模数)变换器16、放大器17、A/D变换器18和19组成,由这些组成输出用于识别作为驾驶员的用户的语音的各种输入数据。

5 具体说,送话器11是例如有方向性的,设置成指向作为驾驶员的用户的方向。因此,在送话器11中主要输入用户的语音。把在送话器11输入的语音变换成音频信号,通过放大器15的放大供给A/D变换器18。在A/D变换器18中,通过放大器15供给的模拟语音信号用规定的取样时钟取样,然后以规定的量化步长量化,因而被变换成数字信号的音频数据。把A/D
10 变换器18输出的音频数据提供给前处理单元20。

CCD摄象机12设置成对用户的口部摄象,把用该CCD摄象机12摄象的用户口部作为视频信号供给A/D变换器16。与在A/D变换器18中的情况相同,A/D变换器16把模拟视频信号变换成图像数据供给前处理单元21。

送话器13例如为无方向性的,它主要输入用户语音之外的声音。即,
15 输入例如引擎声音、汽车装备的收音机和CD播放器等输出的声音、空调器的声音、以及在窗户开着的时候输入外部噪声等。在放大器17和A/D变换器19中,与在放大器15和A/D变换器18中相同地对送话器13输入的声音进行处理,据此作为音频数据提供给前处理单元22。

传感器14可以是测定汽车振动用的加速度传感器或者是测定湿度和温
20 度用的传感器等,其输出供给前处理单元23。加速度传感器的输出表示例如由汽车振动所产生的噪声的程度(大小)。根据测定湿度和温度用的传感器的输出,判定例如是否在下雨,而且,如果在下雨,则测定该雨声大小的程度。

在前处理单元20至23中,分别对供给的数据进行分析,求出表示数
25 据特征的特征参数。

即,前处理单元20和22根据音频数据,以规定时间的语音帧为单位,求出例如音频数据的过零值、在每一规定的频带中的功率电平、线性预测系数、倒频谱(cepsturm)系数等,作为特征参数输出至分类用数据构成单元24和综合参数构成单元26。

30 如图3所示,前处理单元21根据用户口部的图像数据求出例如用户嘴

部的横长 L_1 和纵长 L_2 ，将其比值 L_1/L_2 作为特征参数输出至分类用数据构成单元 24 和综合参数构成单元 26。或者，在前处理单元 21 中，从用户口部的视频数据求出运动向量、边界值、DCT(离散余弦变换)系数等作为特征参数输出至分类用数据构成单元 24 和综合参数构成单元 26。

- 5 在前处理单元 23 中，根据对传感器 14 输出的分析，求出如上所述的例如反映由汽车振动产生的噪声程度(大小)的特征参数或者反映雨声程度的特征参数，这些参数也输出至分类用数据构成单元 24 和综合参数构成单元 26。

- 10 分类用数据构成单元 24 把前处理单元 20 至 23 输出的特征参数的至少一部分作为用于分类的分类用数据输出至分类单元 25。在分类单元 25 中，根据来自分类用数据构成单元 24 的分类用数据的性质，对这些分类用数据按规定类别进行分类。

即，分类单元 25 把预先分配给构成分类用数据的特征参数值的模式(pattern)的值作为该分类用数据的类别供给自适应判定单元 27。

- 15 在这里，若特征参数由例如 A 比特表示，且若分类用数据由 B 个特征参数构成，则构成这样的分类用数据的特征参数值的模式数为 $(2^A)^B$ 。因此如果 A 或 B 很大，类别数就变得巨大，难于使此后的处理迅速进行。

- 20 因此，作为进行分类之前的前处理，对分类用数据进行例如 ADRC(自适应动态范围编码)之类的处理，以减少构成分类用数据的特征参数的比特数。

- 25 即，在 ADRC 处理中，首先从构成分类用数据的 B 个特征参数中检出数值最大的(以下适当地称作最大特征参数)和数值最小的(以下适当地称作最小特征参数)。然后，计算最大特征参数的值 MAX 和最小特征参数的值 MIN 的差值 $DR(=MAX-MIN)$ ，将其作为分类用数据的局部动态范围；根据该动态范围 DR，构成分类用数据的各个特征参数被再量化成比原来分配的比特数 A 少的 C 比特。即，从构成分类用数据的各个特征参数减去最小特征参数的值 MIN，把各减后值除以 $DR/2^C$ 。

- 30 其结果是，构成分类用数据的各个特征参数用 C 比特来表示。因此，在例如设 $C=1$ 的场合，B 个特征参数的模式数成为 $(2^1)^B$ ，与不进行 ADRC 的场合比较，能使模式数变得非常少。

从减少构成分类用数据的特征参数的模式数的观点看, 构成分类用数据的特征参数的数目 B 希望是不怎么大的数值。但是, 特征参数的数目 B 过于小, 也有不能进行恰当分类的危险。因此特征参数的数目 B 希望通过对以上二者平衡后加以确定。

5 另一方面, 综合参数构成单元 26 综合(集合)前处理单元 20 至 23 输出的特征参数的至少一部分而作为综合参数输出至自适应判定单元 27。

在这里, 综合参数所综合和构成的特征参数组可以与构成分类用数据的特征参数组相同, 也可以与构成分类用数据的特征参数组不同。

自适应判定单元 27 由标准参数存储单元 28 和匹配单元 29 组成。当分
10 别从分类单元 25 接收到类别或从综合参数构成单元 26 接收到综合参数时, 自适应判定单元 27 首先从标准参数存储单元 28 选择一登记有与从分类单元 25 接收的类别相对应的标准参数的标准参数表。

即, 标准参数存储单元 28 存储有与类别数目相应的标准参数表, 每个标准参数表登记有由后述学习装置(图 4)通过学习处理获得的一组与例如音
15 素对应的标准参数。在匹配单元 29 中, 从所述每个类别的标准参数表之中选择与分类单元 25 输出的类别相对应的表格。

而且, 匹配单元 29 算出在所选的标准参数表中登记的各个标准参数与来自综合参数构成单元 26 的综合参数之间的例如欧几里德(Euclidian)距离, 把与该欧几里德距离为最小的标准参数相对应的音素作为语音识别结果输
20 出。

因此, 由于有了这种语音识别装置, 不但利用了主要输入用户语音的送话器 11 输出的语音数据, 还利用了从 CCD 摄象机 12 输出的图像数据中得到的例如用户的口部状态和动作、从送话器 13 输出的音频数据和传感器 14 检测出的其它数据, 例如不同类型和频带成分的噪声等, 来识别用户的
25 语音, 所以能提高其识别率。

而且, 为每个类别准备了标准参数表, 并从中选择与输入单元 10 输出的多种数据所确定的类别相对应的标准参数模式, 所以能按照输入单元 10 输出的多种数据而使用对识别用户语音最适当的标准模式表, 其结果是能够进一步提高识别率。

30 图 4 表示进行学习处理的学习装置的构成例, 该学习装置用于求取要登记在图 2 所示的标准参数存储单元 28 中存储的每个类别的标准参数表中

的标准参数。

图 4 中, 输入单元 30(送话器 31、CCD 摄象机 32、送话器 33、传感器 34、放大器 35、A/D 变换器 36、放大器 37、A/D 变换器 38 和 39)、前处理单元 40 至 43、分类用数据构成单元 44、分类单元 45、或综合参数构成单元 46 在结构上分别与在图 2 中的语音识别装置中的输入单元 10(送话器 11、CCD 摄象机 12、送话器 13、传感器 14、放大器 15、A/D 变换器 16、放大器 17、A/D 变换器 18 和 19)、前处理单元 20 至 23、分类用数据构成单元 24、分类单元 25、或综合参数构成单元 26 相同。存储器 47 设有接收从分类单元 45 输出的类别作为地址的地址端(AD)。该存储器 47 存储综合参数构成单元 46 输出的综合参数。

在如上构成的学习装置中, 输入单元 30 输入进行学习用的学习数据。即, 送话器 31 输入例如某个说话人发话的语音。而且, 此时说话人的嘴部用 CCD 摄象机 32 摄象。此外, 送话器 33 输入例如各种汽车的引擎声、用 CD 播放器再现的乐曲、雨声、空调器声、以及其它环境噪声。传感器 34 检测各种振动的程度, 而且, 当送话器 33 输入雨声时, 检测该降雨状态下的温度、湿度等。

然后, 由输入单元 30 接收的学习数据在前处理单元 40 至 43、分类用数据构成单元 44、分类单元 45、或综合参数构成单元 46 中分别进行与在图 2 的前处理单元 20 至 23、分类用数据构成单元 24、分类单元 25、或综合参数构成单元 26 中所进行处理相同的处理。因此, 分类单元 45 向存储器 47 供给类别, 综合参数构成单元 46 也向该存储器供给综合参数。

存储器 47 在与来自分类单元 45 的类别相对应的地址上存储来自综合参数构成单元 46 的综合参数作为标准参数。

以上的处理是对说话人发话的各个音素以及由送话器 33 及传感器 34 输入的各种噪声和数据进行的。

其结果是, 在存储器 47 中, 属于同一类别的综合参数按照地址分类存储。

存储在存储器 47 的各地址中的综合参数(组)保存在图 2 的标准参数存储单元 28 中, 作为标准参数登记为与各类别对应的标准参数表中的存储内容。

由于有了该学习装置, 在无噪声输入送话器 33 和有噪声输入送话器 33

的场合下产生的数据，在分类单元 45 中分类成不同类别。其结果是构成了对有声场合最适合的标准参数表和对无噪声场合最适合的标准参数表。而且，在图 2 的语音识别装置中，由分类单元 25 输出的类别与分类单元 45 输出的类别相同，所以能在有声场合下选择对该场合最适合的标准参数表，而在无噪声的场合下选择对该场合最适合的标准参数表。

而且，在学习装置中，如上所述，对标准参数不仅按照有声场合和无噪声场合而且按照噪声的种类和程度、说话人发话的音素种类和说话人是男性或女性等分类成各种类别。但是怎样对标准参数分类不是重要问题。这是因为，例如，如果语音识别装置输入单元 10 的输入数据与学习装置输入单元 30 的输入数据相同，则在语音识别装置分类单元 25 中的分类结果与在学习装置分类单元 45 中的分类结果相同。因此，在语音识别装置中，可以使用由输入单元 10 输入的数据确定的标准参数或对该输入数据最合适

的标准参数来进行语音识别。

而且，在图 4 所示的学习装置中，能够按照每一个类别对某个音素把一组综合参数存储在存储器 47 中。即，在学习过程中，让一个说话人在种种不同噪声环境下发出同一音素及让多个说话人作这样的发音，会对某一类别的某个音素得到散布在参数空间某种程度范围内的多个综合参数。

具体说，为简单起见，假设综合参数有三个成分 P_1 、 P_2 、 P_3 ，图 5(A) 表示三维的参数空间。如果对某个类别得到的音素“a”和“i”的综合参数进行作图，在参数空间中它们散布在一定的区域内。

虽然这样散布在该区域内的所有点都可以作为该音素的标准参数，但是最好是求出如图 5(B)所示的该区域重心，将此重心作为该音素的标准参数。

图 6 表示图 1 中输入装置 4 内装的语音识别装置第二实施例的组成例。图中与图 2 对应的部分用同一标号表示，在以下适当省略对这些部分的说明。简言之，该第二实施例的语音识别装置设有标准参数存储单元 28_1 至 28_M 来代替标准参数存储单元 28，并且新设立了分类用数据构成单元 51 和分类单元 52，其余基本上与图 2 的语音识别装置相同。

分类用数据构成单元 51 根据输入单元 10 输出的多种数据而构成分类用数据，将其输出至分类单元 52。分类单元 52 对分类用数据构成单元 51 输出的分类用数据进行分类，把作为其分类结果的类别供给前处理单元 20

至 23。

此时，在前处理单元 20 至 23 中，进行与分类单元 52 输出的类别对应的前处理。即，例如，输入到送话器 11 的语音包含较多的元音等有声音时，采用线性预测系数和倒频谱系数等比过零值等更易提取语音的特征，而输入到送话器 11 的语音包含较多的辅音等无声音时，则采用过零值和每个频带的功率电平、辅音的持续时间比线性预测系数和倒频谱系数等更易提取语音的特征。而且，当输入到送话器 13 的噪声电平小时，不必考虑该噪声的影响，但是当噪声电平大时则希望考虑其影响进行语音识别。而且，若用户嘴部很少或没有运动，则不需要考虑其运动向量；如果用户的嘴部产生一定程度的运动，则在语音识别时需考虑其运动向量。而且，当汽车没有或很少振动以及不下雨或下雨很小等等时，不必考虑传感器 14 的输出；在相反的场所，则希望考虑传感器 14 的输出进行语音识别。

所以，在识别语音时最合适的特征参数(为获得较高正确率的识别结果的特征参数)随其它因素而变，更不用说随作为识别对象的语音而变。

另外，在图 6 的语音识别装置中，根据输入单元 10 输出的数据来构成分类用数据，对其进行分类，然后分别在前处理单元 20 至 23 中求出对每种类别的最适当特征参数。

因此，在图 6 的实施例 1 中，在自适应判定单元 27(匹配单元 29)中计算距离用的参数空间随分类单元 52 的分类结果而变化。即，在自适应判定单元 27 中计算在与分类单元 52 输出的类别相对应的参数空间中的距离，根据该距离来输出语音识别结果。

假定在分类单元 52 中，把分类用数据构成单元 51 输出的分类用数据均分成 M 个类别。

而且，在前处理单元 20 至 23 中，除了确定与分类单元 52 输出的类别相对应的特征参数之外，还可以改变特征参数的阶数(degree)(例如第 8 阶或第 12 阶的线性预测系数)，或者可以取消特征参数的输出(例如，当汽车在寂静场所停车的时候，可不必考虑送话器 13 和传感器 14 的输出，因此使前处理器 22 和 23 处于休眠状态)等等。

分类单元 52 输出的类别，除了供给前处理单元 20 至 23 外，也供给自适应判定单元 27。如上所述，自适应判定单元 27 有 M 个标准参数存储单元 28₁ 至 28_M，其中分别存储着分别与在分类单元 52 输出的 M 个类别对应

的参数空间中的标准参数。

而且，在标准参数存储单元 $28_m(m = 1, 2, \dots, M)$ 中，与在图 2 的标准参数存储单元 28 中的情况相同，存储着分类单元 25 输出的每个类别的标准参数表。

- 5 存储在标准参数存储单元 28_1 至 28_M 中的标准参数表，可以用后述的学习装置(图 7)通过学习求出。

一旦从分类单元 52 收到类别，自适应判定单元 27 就从标准参数存储单元 28_1 至 28_M 中选择与该类别对应的存储单元(以下把从标准参数存储单元 28_1 至 28_M 中选出的存储单元适当地记为标准参数存储单元 28_{ms})。

- 10 另一方面，前处理单元 20 至 23 输出的特征参数通过分类用数据构成单元 24 供给分类单元 25 以进行分类，然后作为该分类结果的类别则供给自适应判定单元 27。而且，前处理单元 20 至 23 输出的特征参数也供给综合参数构成单元 26，在该单元内形成综合参数。根据在由自适应判定单元 27 选出的标准参数存储单元 28_{ms} 中存储的标准参数表中登记的标准参数的
15 同一参数空间中的参数，由综合参数构成单元 26 构成综合参数。

- 在综合参数构成单元 26 中得到的综合参数供给自适应判定单元 27。在自适应判定单元 27 中，按照分类单元 25 输出的类别，来选择标准参数存储单元 28_{ms} 存储的一个标准参数表。在匹配单元 29 中，算出登记在所选标准参数表中各标准参数的每一个标准参数与来自综合参数构成单元 26 的综合参数之间的距离。把与综合参数的距离为最小的标准参数对应的音素作
20 为语音识别结果来输出。

如上所述，进行与分类单元 52 输出的类别相对应的前处理，得到了在各种场合下最适当的特征参数，因此能根据对各场合最合适的参数空间中的距离得到精度更高的语音识别结果。

- 25 图 7 表示进行学习处理的学习装置的组成例，该学习装置用于求出要登记在图 6 的各个标准参数存储单元 28_1 至 28_M 中存储的每个类别的标准参数表中的标准参数。

- 在该图中，与图 4 对应的部分用相同的标号表示，在下文中适当省略对其的说明。该学习装置设有存储器 47_1 至 47_M 来取代存储器 47，而且新
30 设立了分类用数据构成单元 61、分类单元 62 和选择器 63，其它组成基本上和图 4 的学习装置相同。

在分类用数据构成单元 61 或分类单元 62 中, 分别进行与图 6 的分类用数据构成单元 51 或分类单元 52 中所进行处理的相同处理, 作为分类处理结果的类别供给前处理单元 40 至 43 和选择器 63。在前处理单元 40 至 43 中进行与图 6 的前处理单元 20 至 23 中所进行处理的相同处理, 从而得到并输出与分类单元 62 输出的类别对应的最适合的特征参数。

而且, 在分类用数据构成单元 44、分类单元 45、或综合参数构成单元 46 中, 分别进行与图 6 的分类用数据构成单元 24、分类单元 25、或综合参数构成单元 26 的情况同样的处理, 其结果是分别从分类单元 45 输出类别, 从综合参数构成单元 46 输出综合参数。

另一方面, 选择器 63 随着分类单元 62 输出的类别把选择信号输出至存储器 47_1 至 47_M 中任一存储器的片选(CS)端。即, 由此在存储器 47_1 至 47_M 中选择与分类单元 62 输出的类别相对应的存储器。

又, 分类单元 45 输出的类别供给存储器 47_1 至 47_M 的地址(AD)端, 因此, 按照分类单元 62 输出类别选择的与分类单元 45 输出类别对应的存储器地址, 综合参数构成单元 46 输出的各个综合参数被存储在存储器 47_1 至 47_M 中。

如上所述, 分别存储在标准参数存储器 47_1 至 47_M 中各个地址中的综合参数(组)被分别存储在图 6 的标准参数存储器 28_1 至 28_M 中相应的一个存储器, 作为登记在与各个类别(分类单元 25 输出的类别)对应的标准参数表中的标准参数。

在此情况下, 也可以如图 5 所说明的, 在图 6 的各个标准参数存储单元 28_1 至 28_M 中存储散布参数空间内的一组综合参数的重心之类, 作为标准参数。

图 8 表示图 1 输入装置 4 内装的语音识别装置第三实施例的组成例。在图 8 中, 对与图 6 对应的部分给予相同标号, 在以下省略对这些部分的说明。此实施例的语音识别装置新设立前处理单元 71 至 74, 前处理单元 71 至 74 的输出取代了前处理单元 20 至 23 的输出而供给综合参数构成单元 26, 其余的组成与图 6 的语音识别装置基本上相同。

给予前处理单元 71 至 74 的输入与给予前处理器 20 至 23 的输入相同。而且, 也把分类单元 52 输出的类别供给前处理器 71 至 74。

在前处理单元 71 至 74 中, 进行与分类单元 52 输出的类别相对应的前

处理, 因此把最适合的特征参数输出至综合参数构成单元 26。在前处理单元 71 至 74 的每一个中, 进行与前处理单元 20 至 23 的每一个中基本上不同的前处理。即, 根据前处理单元 20 至 23 的输出, 最终得到分类单元 25 输出的类别, 与此相反, 根据前处理单元 71 至 74 的输出, 得到的是综合参数构成单元 26 输出的综合参数。因此, 在前处理单元 20 至 23 中, 与分类单元 52 输出的类别对应地求出适用于在分类单元 25 中进行分类的最合适的特征参数, 另一方面, 在前处理单元 71 至 74 中, 则与分类单元 52 输出的类别对应地求出对于构成最适用于语音识别的综合参数来说最适合的特征参数。

10 图 9 表示进行学习处理的学习装置的组成例, 该装置用于求取要登记在标准参数存储单元 28_1 至 28_M 中存储的每个类别的标准参数表中的标准参数。

在图 9 中, 与图 7 对应的部分用相同的标号表示, 在下文中适当省略对其的说明。该学习装置新设立前处理单元 81 至 84, 前处理单元 81 至 84 15 的输出取代前处理单元 40 至 43 的输出而供给综合参数构成单元 46, 其它组成基本上和图 7 的学习装置相同。

与图 8 的前处理单元 20 至 23 的情况相同, 在前处理单元 40 至 43 中与分类单元 62 输出的类别相对应地求出用于在分类单元 45 中进行分类的最合适的特征参数, 而且, 与图 8 的前处理单元 71 至 74 的情况相同, 前 20 处理单元 81 至 84 与分类单元 62 输出的类别相对应地求出对于构成最适用于语音识别的综合参数来说最合适的特征参数。

虽然可以把通过利用图 9 的学习装置学习处理得到的综合参数存储在图 8 的标准参数存储单元 28_1 至 28_M 中, 但是, 也可以不全部存储通过学习得到的综合参数。如图 5 所说明的, 可以存储在参数空间中散布的一组综合 25 参数的重心之类作为标准参数。

以上对应用本发明的语音识别装置作了说明, 但是本发明也可应用于除语音之外的例如图像和文字、人物等其它对象的识别装置。

而且, 在本实施例中, 除了使用作为识别对象的用户的语音之外, 还使用 CCD 摄象机 32、送话器 33 和传感器 34 的输出进行识别, 但是不局限 30 于这些输出。

而且, 在图 2 的实施例中, 根据前处理单元 20 至 23 的输出数据而构

成分类用数据，并在分类单元 25 中把该分类用数据进行分类。但是也可以直接从例如输入单元 10 的输出数据来构成分类用数据，用于在分类单元 25 中分类。

而且，虽然在图 6 至 9 的实施例 5 中向前处理单元 20 至 23、40 至 43、71 至 73、81 至 83 提供类别并进行与该类别相对应的前处理，但是也可以向前处理单元 20 至 23、40 和 43、71 至 73、81 至 83 供给与类别对应的函数，以便利用该函数进行数学运算来进行与该类别对应的前处理。

而且，为了简化本实施例的说明，在匹配单元 29 中根据在某个所采用参数空间的综合参数与标准参数之间的距离而得到语音识别结果，但是在匹配单元 29 中也可以根据按时间序列规定的综合参数与标准参数之间的距离和根据此种时间序列的出现概率来求得语音识别结果。此外，在匹配单元 29 中，也可以对分类单元 25 和 52 输出的每个类别用不同的语音识别算法来进行语音识别。

而且，图 2、图 6 和图 8 中说明的语音识别装置以及图 4、图 7 和图 9 中说明的学习装置不仅可用硬件来实现，也可以作为具有 CPU(中央处理单元)和存储器等的微处理器的应用以软件方式实现。

按照权利要求 1 所述的识别装置和权利要求 5 所述的识别方法，根据多种输入数据的性质把这些数据按规定类别分类，同时构成对多种输入数据综合的综合参数，根据登记了与分类结果得到的每个类别相对应的标准参数的标准参数表和根据综合参数而对识别对象进行识别。因此，使用了最适合于每一种场合的表，所以能提高识别性能。

按照权利要求 6 所述的学习装置和权利要求 9 所述的学习方法，根据多种输入数据的性质把这些数据按规定类别分类，同时构成对多种输入数据综合的综合参数，把该综合数据按照由分类结果得到的每个类别来分类。因此，对于每一种情况，可得到最适合于识别用的参数。

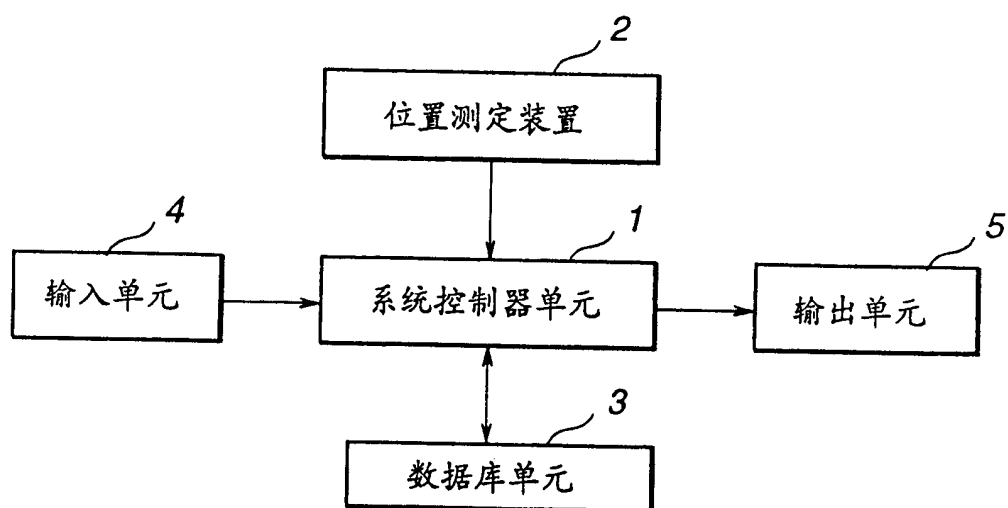


图 1

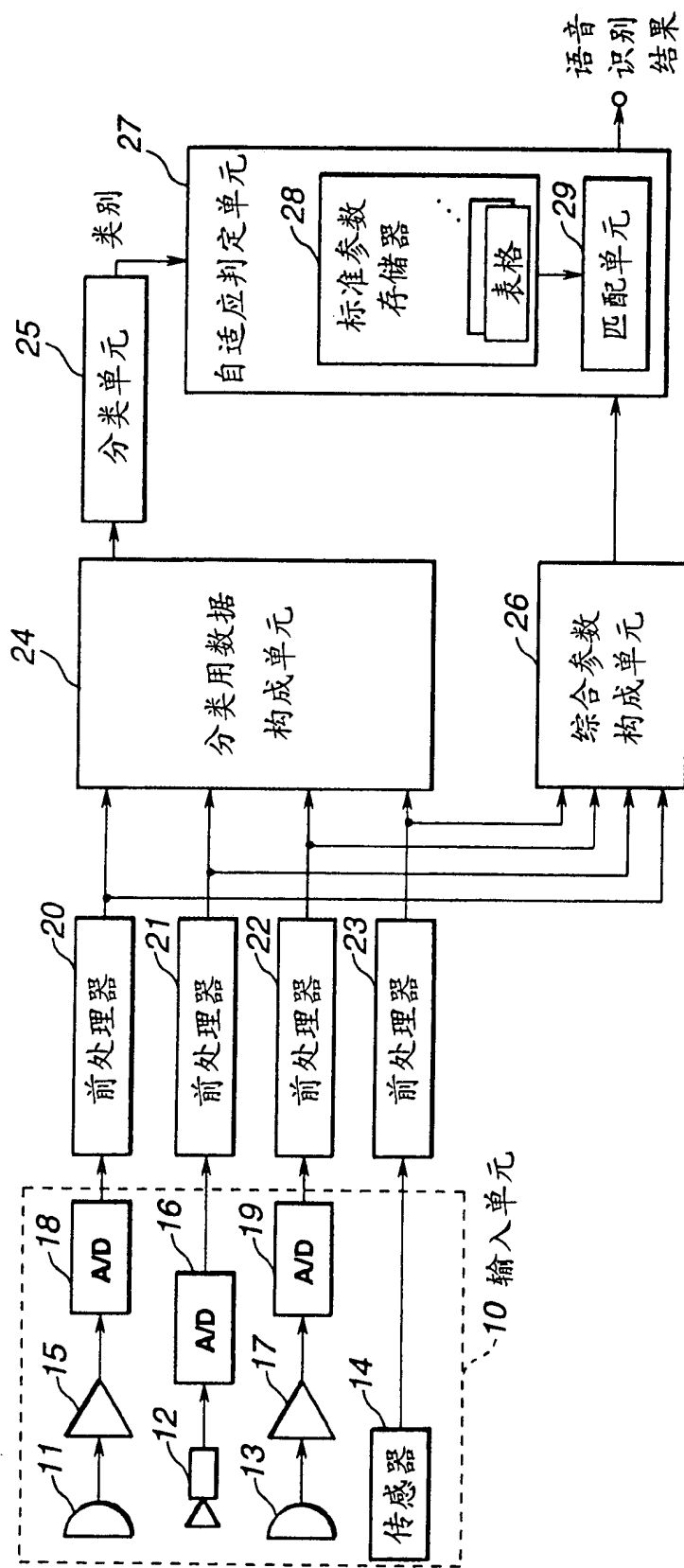


图 2

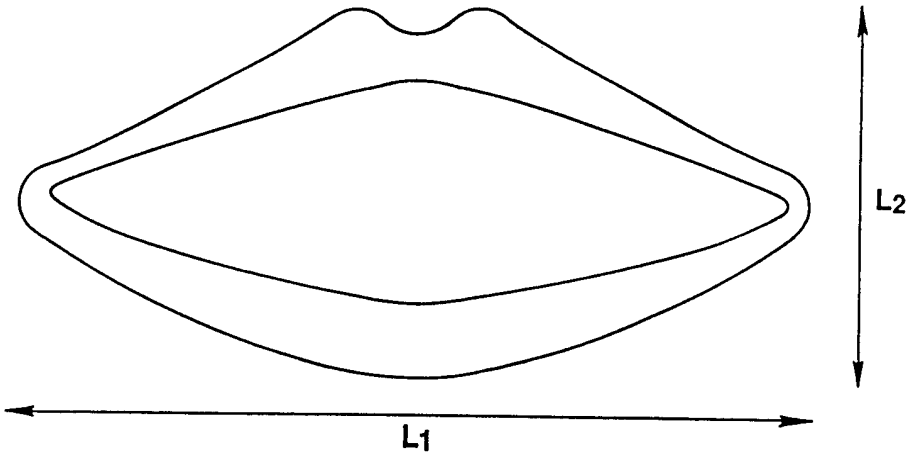


图 3

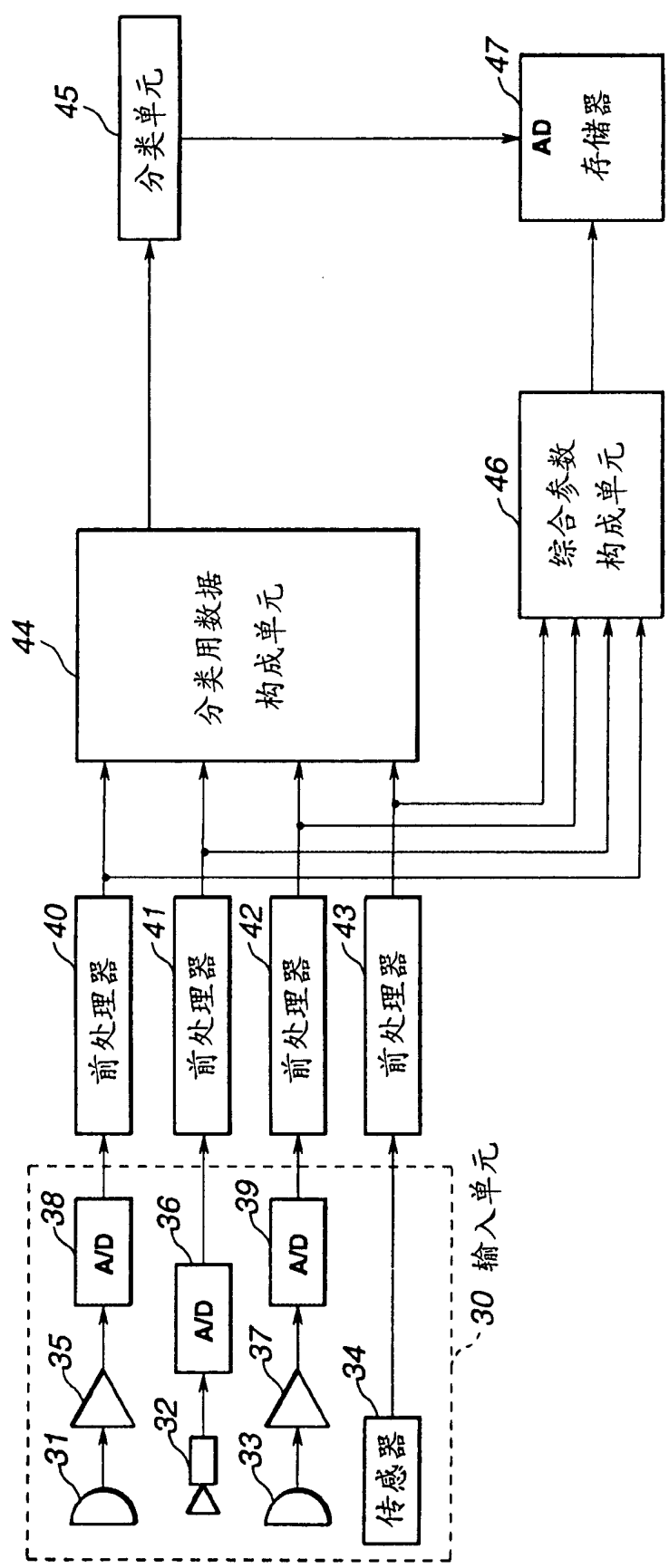


图 4

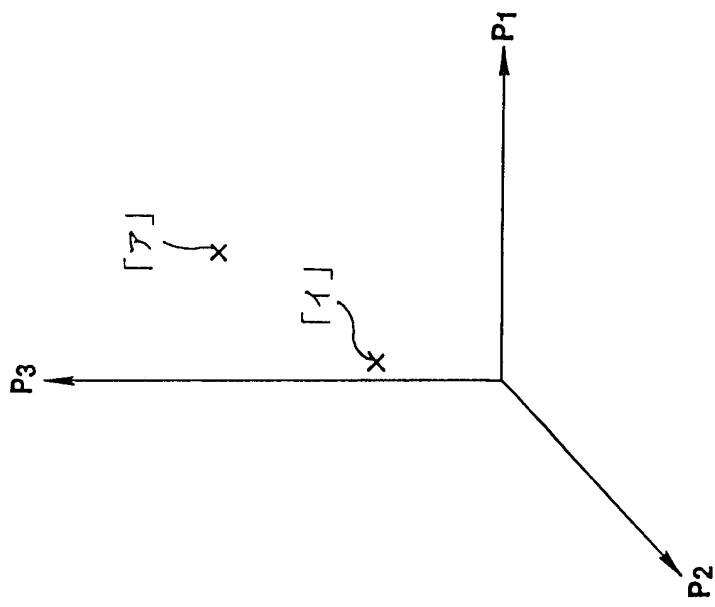
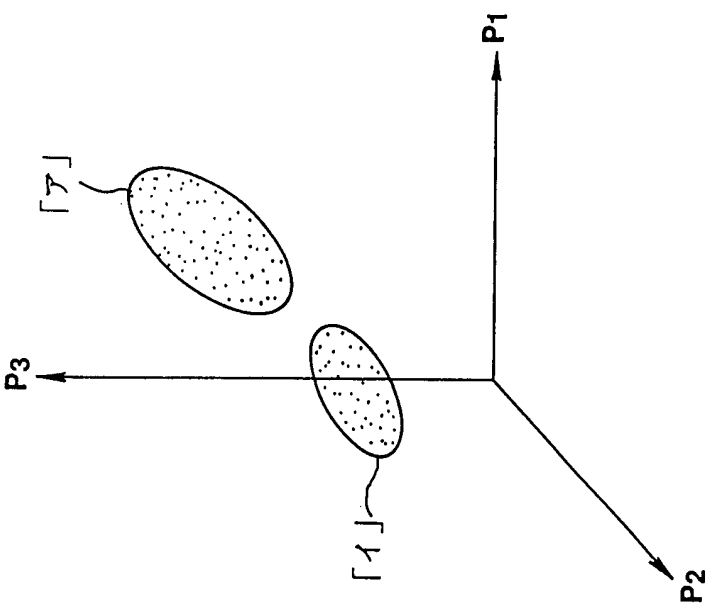


图 5



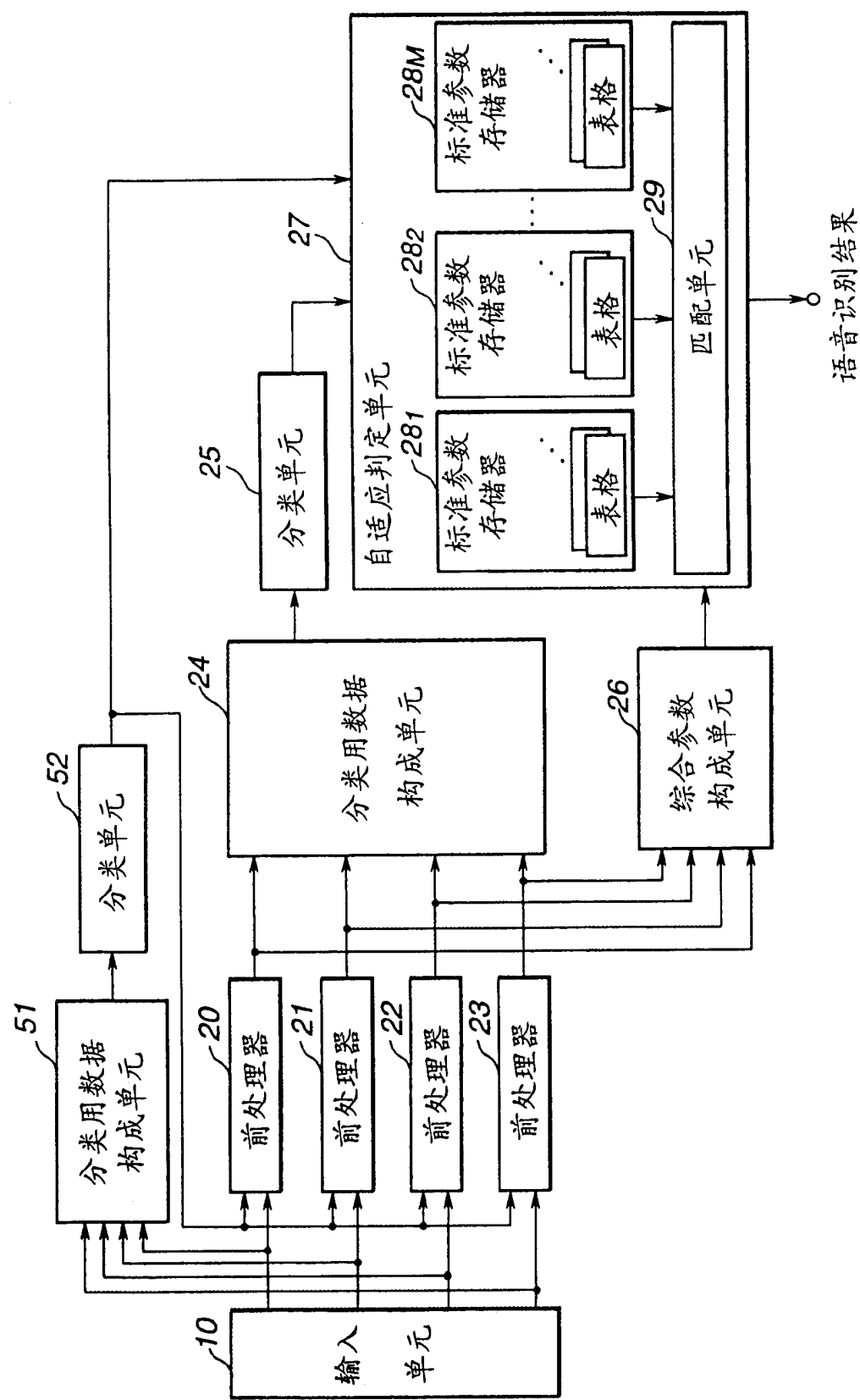


图 6

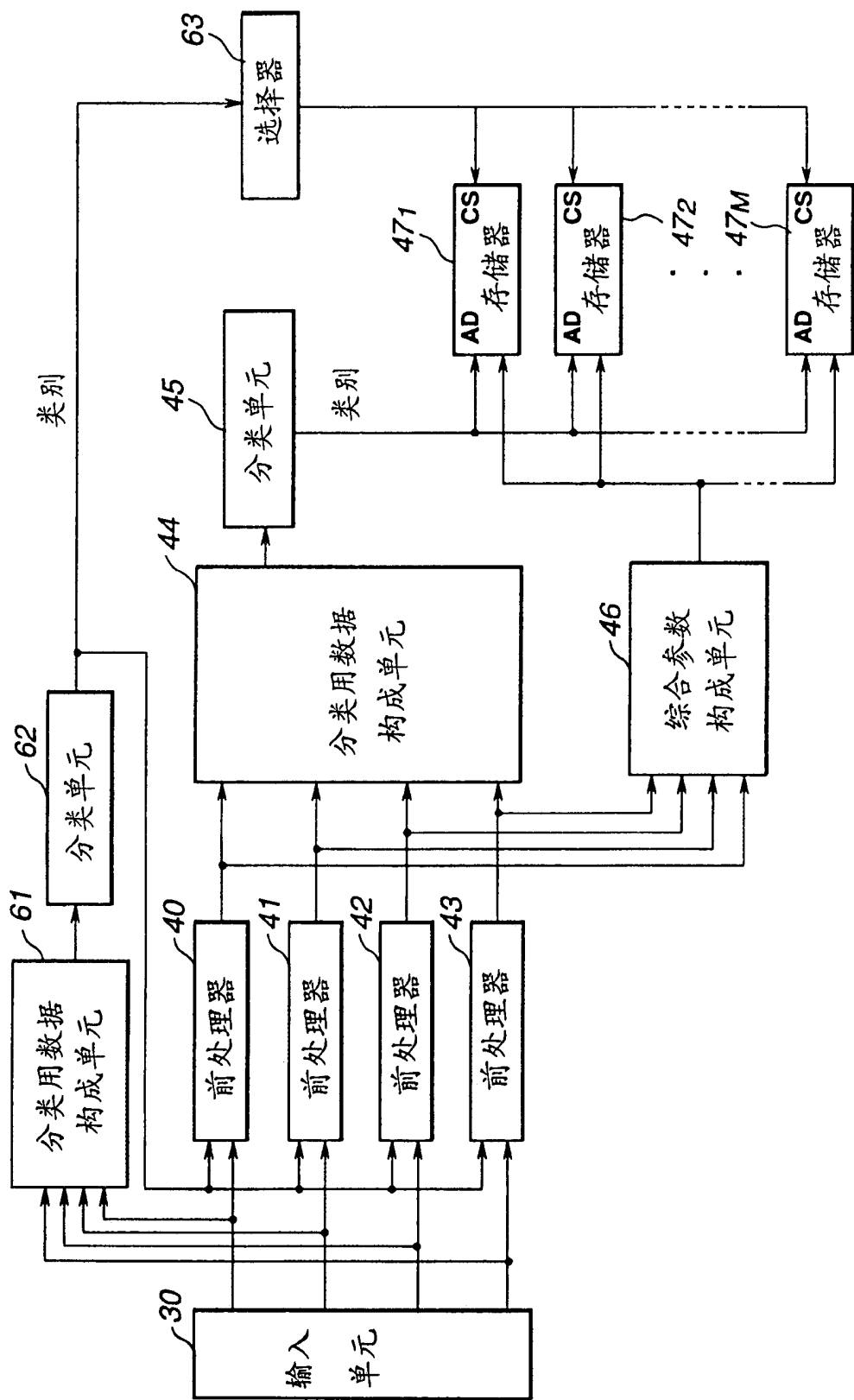
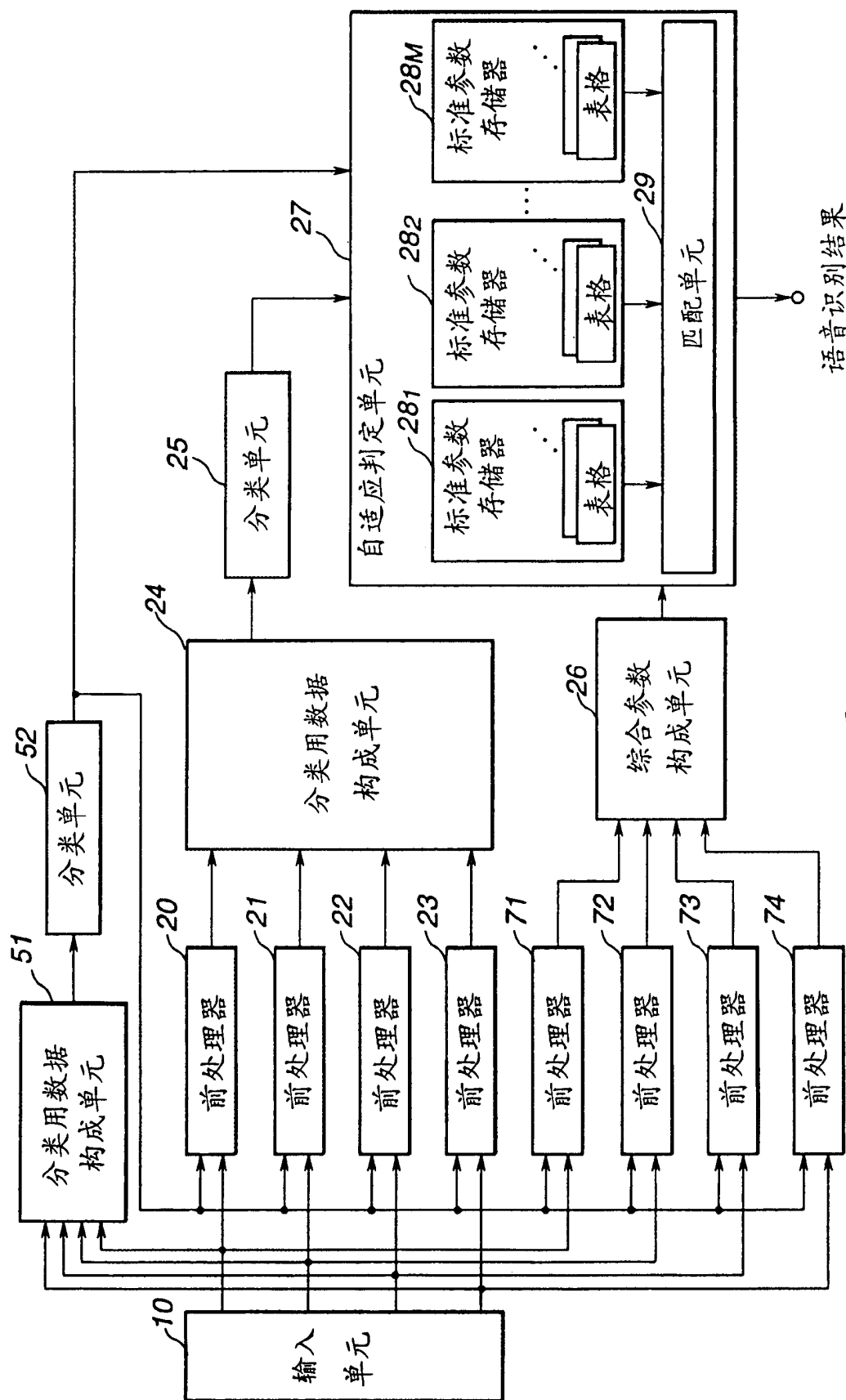


图 7



语音识别结果

图 8

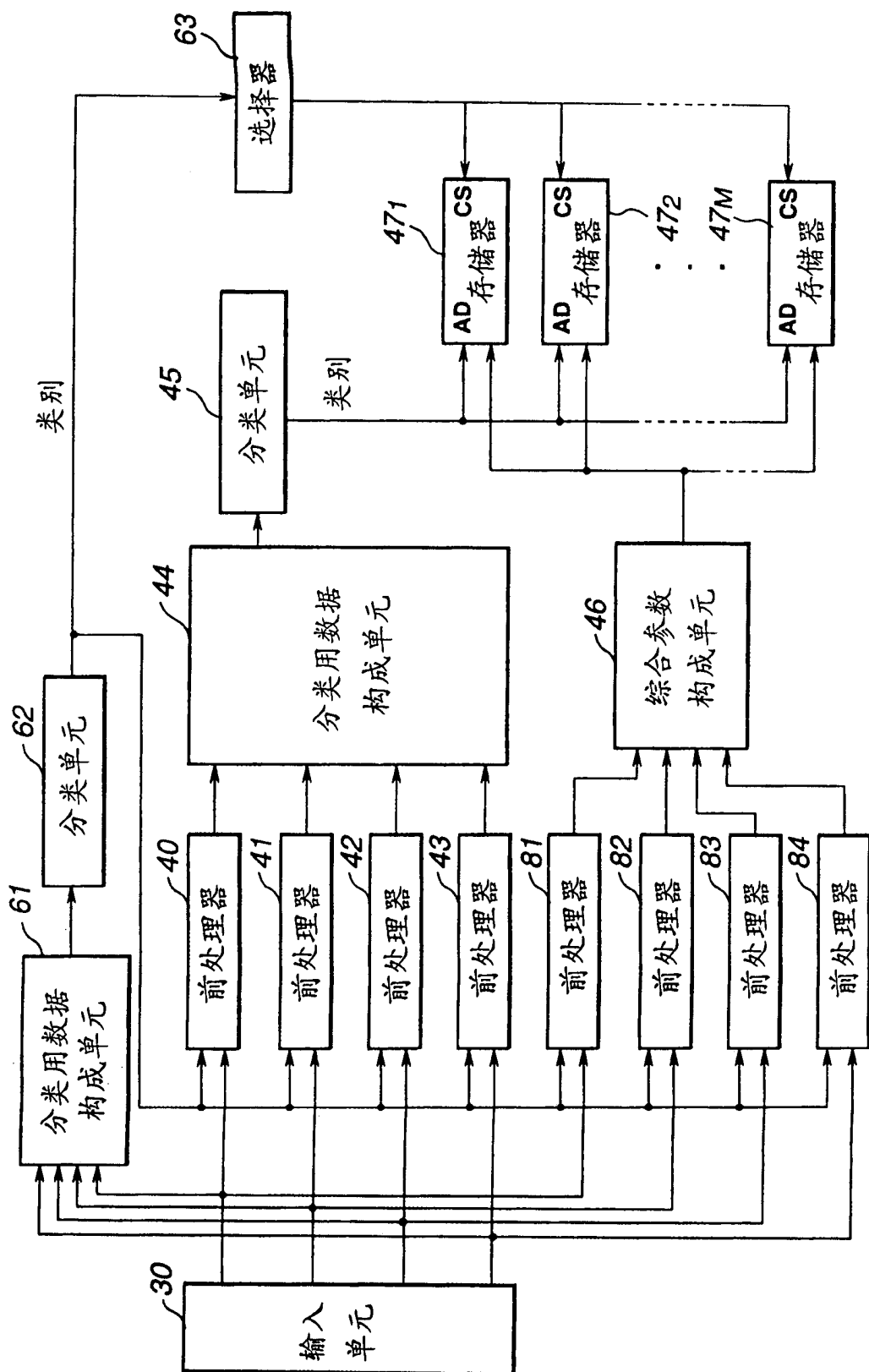


图 9