



[12] 发明专利申请公开说明书

[21] 申请号 03122309.5

[43] 公开日 2003 年 11 月 5 日

[11] 公开号 CN 1453767A

[22] 申请日 2003.4.23 [21] 申请号 03122309.5

[30] 优先权

[32] 2002. 4. 26 [33] JP [31] 126939/2002

[32] 2002. 5. 17 [33] JP [31] 142998/2002

[71] 申请人 日本先锋公司

地址 日本东京

[72] 发明人 外山聪一

[74] 专利代理机构 北京三友知识产权代理有限公司

司

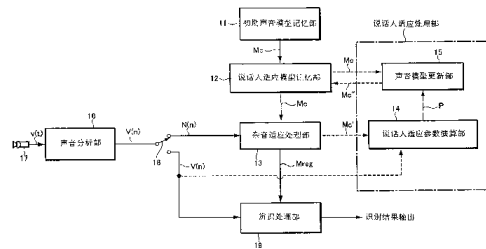
代理人 李 辉

权利要求书 4 页 说明书 31 页 附图 8 页

[54] 发明名称 语音识别装置以及语音识别方法

[57] 摘要

一种语音识别装置以及语音识别方法，借由较高的精度以及较少的处理量对语音识别的结果作正误判断。借由对照声音模型 HMMsb 与语音的特征矢量 $V(n)$ 求取表示最大优度的声音模型的识别结果 RCG 与表示最大优度值的第一记录 FSCR，与表示其次优度值的第二记录 SSCR，而借由将以前述第一记录 FSCR 及第二记录 SSCR 为依据的评价值 $FSCRX(FSCR - SSCR)$ 与预先设定的基值 THD 比较，来判断识别结果为正或为误。识别结果 RCG 若判断为正确时则对声音模型 HMMsb 实行说话人适应处理，而识别结果 RCG 若判断为误，则不对声音模型 HMMsb 进行说话人适应处理，借此方式以提高说话人适应处理的精度等。



1、一种语音识别装置，其对声音模型借由语音的特征矢量实施说话人适应处理；其具备有：

- 5 语音识别装置，借由对照上述声音模型与语音的特征矢量而输出表示得到最大优度（相似度）的声音模型的识别结果，与表示上述最大优度的值的第一记录，与表示其次的优度值的第二记录；

判断装置，将依据上述第一记录与第二记录的评价值与预先设定的基值(阈值)作比较，在上述评价值相对于基值呈一定关系的情况下则判
10 断前述识别结果为正确；

说话人适应处理装置，借由上述判断装置判断上述识别结果为正确时对前述声音模型实施说话人适应处理。

- 2、根据权利要求1所述的语音识别装置，其中前述判断装置在上述评价值相对于基值不呈一定关系的情况下判断上述识别结果为误，而上述说话人适应处理装置在判断上述识别结果为误时不对前述声音模型实施说话人适应处理。
15

3、根据权利要求1所述的语音识别装置，其中前述评价值根据上述第一记录与第二记录的差值进行计算。

- 4、根据权利要求2所述的语音识别装置，其还具备一装置，在前述
20 判断装置判断前述识别结果为错误时则禁止上述识别结果的输出，并且作出表示上述识别结果为错误的信息提示。

5、一种语音识别方法，根据语音的特征矢量对声音模型实施说话人适应处理；其具备有：

- 第一步骤，对照上述声音模型与语音的特征矢量而求取表示得到最大优度的声音模型的识别结果，与表示前述最大优度的值的第一记录，
25 及表示其次优度值的第二记录；

第二步骤，将依据前述第一记录与第二记录得出的评价值与预先设定的基值作比较，在上述评价值相对于基值呈一定关系的情况下判断上述识别结果为正确；

第三步骤，在上述第二步骤中，在判断前述识别结果为正确时则对前述声音模型实施说话人适应处理。

6、根据权利要求5所述的语音识别方法，其在上述第二步骤中在上述评价价值相对于前述基值不呈一定关系的情况下则判断上述识别结果为误，而在上述第三步骤中在判断上述识别结果为误时则不对上述声音模型实施说话人适应处理。

7、根据权利要求5所述的语音识别方法，其中前述评价价值借由上述第一记录与第二记录的差值作计算。

8、根据权利要求6所述的语音识别方法，其在前述第二步骤中在判断前述识别结果为错误时则禁止上述识别结果的输出，同时还作出表示上述识别结果为错误的信息提示。

9、一种语音识别装置，具备有：

记忆装置，具有初期语音模型；

一杂音适应装置，借由在对上述记忆装置的初期语音模型实行说话人适应处理时的背景杂音，实施以杂音适应处理来产生杂音适应模型；

说话人适应参数计算装置，对上述杂音适应处置所产生的前述杂音适应模型借由上述说话人适应处理时所说话的语音实行说话人适应计算，从而计算出将上述杂音适应模型变换为杂音叠加说话人适应模型用的说话人适应参数；

语音模型更新装置，在对上述记忆装置的初期语音模型借由上述说话人适应处理参数实施说话人适应处理以产生说话人适应模型，而将该说话人适应模型取代前述初期语音模型从而更新记忆于上述记忆装置。

10、根据权利要求9所述的语音识别装置，其还具备有在语音识别时实行语音识别处理的识别处理装置，并且上述杂音适应装置借由上述语音识别时的非说话期间的背景杂音对更新记忆于上述记忆装置的上述说话人适应模型实施杂音适应处理，从而产生经杂音适应处理的说话人适应模型，并且将杂音适应处理过的说话人适应模型作为识别语音用的声音模型供给至上述语音识别装置。

11、一种语音识别装置，具备有：

记忆装置，具有初期语音模型；杂音适应装置，借由语音识别时的非说话期间的背景杂音对上述记忆装置的初期语音模型实施杂音适应处理而生成杂音适应模型；

5 识别处理装置，对照上述语音识别时的说话期间所说话的应该语音识别的语音与上述杂音适应装置所生成的上述杂音适应模型而实行语音识别；

说话人适应参数计算装置，对前述杂音适应装置所生成的上述杂音适应模型借由上述应该语音识别的语音实行说话人适应计算，而算出将上述杂音适应模型变换成杂音叠加说话人适应模型用的说话人适应参
10 数；

语音模型更新装置，对上述记忆装置的初期语音模型借由上述说话人适应参数实施说话人适应处理以产生说话人适应模型，并将该说话人适应模型取代上述初期语音模型而更新记忆于上述记忆装置中。

12、根据权利要求 11 所述的语音识别装置，其中前述说话人适应参数计算装置与语音模型更新装置在前述识别处理装置的识别结果的可靠
15 度较高时，生成前述说话人适应模型而取代上述初期语音模型从而更新记忆于上述记忆装置中。

13、一种语音识别方法，包括：

杂音适应处理步骤；对记忆于记忆装置的初期语音模型借由说话人
20 适应处理时的背景杂音实施杂音适应处理而产生杂音适应模型；

说话人适应参数计算处理步骤，对上述杂音适应处理步骤所产生的上述杂音适应模型借由上述说话人适应处理时所说话的语音实行说话人适应计算，而计算出将上述杂音适应模型变换成杂音叠加说话人适应模型用的说话人适应参数；

25 语音模型更新处理步骤，对上述记忆装置的初期语音模型借由上述说话人适应参数实施说话人适应处理而产生说话人适应模型，而以该说话人适应模型取代上述初期语音模型从而更新记忆于上述记忆装置。

14、根据权利要求 13 所述的语音识别方法，其在上述杂音适应处理步骤中还具备有语音识别处理步骤，借由在语音识别时的非说话期间的

背景杂音，对更新记忆于上述记忆装置的上述说话人适应模型实施杂音适应处理，从而生成实施过杂音适应处理的说话人适应模型，而借由对照该实施过杂音适应处理的说话人适应模型与上述语音识别时的说话期间的应该语音识别的语音来实行语音识别。

5 15、一种语音识别方法，包括：

杂音适应处理步骤，借由语音识别时的非说话期间的背景杂音对记忆于记忆装置的初期语音模型实施杂音适应处理而产生杂音适应模型；

识别处理步骤，对照上述语音识别时的说话期间所说话的应该语音识别的语音与上述杂音适应处理步骤所生成的上述杂音适应模型而实行
10 语音识别；

说话人适应参数计算处理步骤，对上述杂音适应处理步骤所生成的上述杂音适应模型借由上述应该语音识别的语音实行说话人适应计算，从而算出将上述杂音适应模型变换成杂音叠加说话人适应模型用的说话人适应参数；

15 语音模型更新处理步骤，对上述记忆装置的初期语音模型以上述说话人适应参数实施说话人适应处理而产生说话人适应模型，而将该说话人适应模型取代上述初期语音模型以更新记忆于上述记忆装置。

16、根据权利要求 15 所述的语音识别方法，其中上述说话人适应参数计算处理步骤与语音模型更新处理步骤在上述识别处理步骤的识别结果
20 的可靠度较高的情况下产生上述说话人适应模型，从而取代前述初期语音模型而更新记忆于上述记忆装置中。

语音识别装置以及语音识别方法

5 技术领域

本发明涉及一种借由说话人适应处理实行语音识别的语音识别装置以及语音识别方法。

而且，本发明还涉及一种使用实施过说话人适应处理的语音模型实行语音识别的语音识别装置以及语音识别方法。

10

背景技术

语音识别很困难的主要原因，一般被认为是由于说话人的语音受发音器官及语音习惯等个人差异影响。

以往，对于这些存在个人差异影响的语音，作为常用的语音识别算法，主要使用 MLLR(Maximum Likelihood Linear Regression)与 MAP(Maximum a Posteriori)等说话人适应处理来提高对于特定说话人语音的识别率。

这种说话人适应处理方法，利用许多非特定人的语音预先形成以音素或者单词为单位的许多声音模型(初期的声音模型)，然后借由应作说话人适应处理的说话人语音的特征量，对这些声音模型进行说话人适应处理。

但是，为了实行适当的说话人适应处理，必须参照语音识别结果，对音素或单词单位的声音模型实施说话人适应处理。

也就是，以具体的事例来说明，当说话人讲出“东京”时，只有在该语音被正确识别为“东京”时，才需要将“东京”的声音模型作说话人适应处理，若根据错误识别的结果将“东京”的声音模型作说话人适应处理，则变成实行错误的说话人适应处理之情况。

假设说话人讲出“东京”时，将其语音误识别为“京都”，则当对“京都”的声音模型以“东京”的语音特征量实行说话人适应处理时，无法

实施适当的说话人适应处理，从而会降低说话人适应处理后的语音识别率。

因此，作为实行说话人适应处理时的前提条件，准确的判断语音与语音识别的结果是否一致，也就是准确的判断语音识别的结果是否正确
5 很重要。

但是，按照现有的说话人适应处理方法，因为判断语音识别结果是否正确的工作量非常庞大，因此开发一种不会使说话人感到压力，可迅速并且适当的实行说话人适应处理的新颖的判断处理方法成为一个很重要的课题。

10 此外，在语音识别的技术领域中，存在使用大量语音数据库实行学习的非特定人语音模型实行语音识别的技术。

但是，这种非特定人语音模型，因为是根据许多非特定人的语音资料进行学习，所以在识别标准语音说话人的语音的情况下可得到较高的识别性能，然而在识别具有特征语音说话人的语音的情况下，存在不
15 一定能得到较高识别性能的问题。

因此，本发明有借由各个说话人的语音将非特定人语音模型作说话人适应处理，而借由使用该经说话人适应处理的声音模型对每个说话人实行适当的语音识别的说话人适应处理方法。

现有的说话人适应处理方法，使用大量的语音数据库预先生成音素
20 等次单词单位的非特定人语音模型(以下称为“初期语音模型”)，而在开始实际的语音识别之前的前期处理阶段对初期语音模型实施说话人适应处理。也就是，在前期处理阶段的时候使说话人说话，然后根据语音特征矢量系列将初期语音模型作说话人适应处理，生成考虑到说话人个人差异的说话人适应模型。

25 还有，在实际的语音识别时，对照说话人所发出的应识别的语音的特征矢量系列与由上述的说话人适应模型构成的系列，就可得对说话人适应模型系列作语音识别的最优结果。

但是，对于初期语音模型实行说话人适应处理时，在说话人的语音上将叠加以说话环境下的背景杂音。

因此，应用现有的说话人适应处理方法的情况下，除说话人的语音之外，还需借由叠加有背景杂音的语音(也就是背景杂音叠加语音)的特征矢量系列作说话人适应处理，从而有时会难以产生高精度的说话人适应模型。

- 5 特别是在使用杂音较多环境下的语音进行说话人适应处理时，将强烈的受到背景杂音的影响，而产生难以生成能适当反映说话人特征的说话人适应模型之情形。

而且，实际上，借由利用现有的说话人适应处理方法作说话人适应处理的上述说话人适应模型实行语音识别时，如果识别时环境下的背景杂音与已述的说话人适应处理时的背景杂音不同，就会无法充分获得实行说话人适应处理所产生的效果，也就是无法改善语音识别率。

发明内容

为解决前述问题，本发明的目的在于提供一种可迅速并且适当的实行说话人适应处理的语音识别装置以及语音识别方法。

还有，本发明的另一目的在于提供一种对于背景杂音的影响可实行强制的说话人适应处理的语音识别装置以及语音识别方法。

为实现上述目的，本发明提供一种语音识别装置，其对声音模型借由语音的特征矢量实施说话人自适应；其具备有：语音识别装置，借由对照上述声音模型与语音的特征矢量而输出表示得到最大优度(相似度)的声音模型的识别结果，与表示上述最大优度的值的第一记录，与表示其次的优度值的第二记录；判断装置，将依据上述第一记录与第二记录的评价值与预先设定的基值(阈值)作比较，在上述评价值相对于基值呈一定关系的情况下则判断前述识别结果为正确；说话人适应处理装置，借由上述判断装置判断上述识别结果为正确时对前述声音模型实施说话人适应处理。

另外，本发明提供一种语音识别方法，根据语音的特征矢量对声音模型实施说话人适应处理；包括：第一步骤，对照上述声音模型与语音的特征矢量而求取表示得到最大优度的声音模型的识别结果，与表示前

述最大优度的值的第一记录，及表示其次优度值的第二记录；第二步，
将依据前述第一记录与第二记录得出的评价值与预先设定的基值作比
较，在上述评价值相对于基值呈一定关系的情况下判断上述识别结果为
正确；第三步，在上述第二步中，在判断前述识别结果为正确时则
5 对前述声音模型实施说话人适应处理。

所述的语音识别装置与语音识别方法，其中，在说话人适应处理的
过程中，借由对照声音模型与语音的特征矢量求取表示得到最优声音模
型的识别结果与表示最优值的第一记录以及表示其次优度值的第二记
录，然后再将依据第一记录与第二记录的评价值与预设的基值比较。随
10 后，在评价值相对于基值呈一定关系时，则判断为识别结果正确，对声
音模型实施说话人适应处理。借此，可根据正确的识别结果实施适当的
说话人适应处理。

另外，所述语音识别装置，其中，在上述评价值相对于基值为不呈
一定关系的情况下，则上述判断装置判断上述识别结果为误，而上述说
15 话人适应处理装置在判断上述识别结果为误时不对前述声音模型实施说
话人适应处理。

另外，所述语音识别方法，其中，在上述第二步中在上述评价值
相对于前述基值不呈一定关系的情况下则判断上述识别结果为误，而在
上述第三步中在判断上述识别结果为误时则不对上述声音模型实施说
20 话人适应处理。

所述的语音识别装置与语音识别方法，其中，当识别结果判断为误
时则不对声音模型实施说话人适应处理。借此，借由不根据错误的识别
结果实行说话人适应处理可防止说话人适应处理后的语音识别精度的变
低。

25 所述的语音识别装置，其中前述评价值根据上述第一记录与第二记
录的差值进行演算。

所述的语音识别方法，其中前述评价值根据上述第一记录与第二记
录的差值进行演算。

所述的语音识别装置与语音识别方法，借由以第一记录与第二记录

的差值计算评价值，而在该计算所得的评价值相对于基值呈一定关系时则判断识别结果为正确，另外，其评价值相对于基值不呈一定关系时则判断为识别结果为错误，由此可提高判断时的判断精度。

另外，所述语音识别装置，其还具备一装置，为在前述判断装置判断前述识别结果为错误时则禁止上述识别结果的输出，并且实行上述识别结果为错误的信息提示。

另外，所述语音识别方法，在前述第二步骤中在判断前述识别结果为错误时则禁止上述识别结果的输出同时还实行上述识别结果为错误的信息提示。

10 所述的语音识别装置与语音识别方法，可对使用者提示是否实行了适当的说话人适应处理等有效信息。

此外，本发明提供一种语音识别装置，具备有：记忆装置，具有初期语音模型；一杂音适应装置，借由在对上述记忆装置的初期语音模型实行说话人适应处理时的背景杂音，实施以杂音适应处理来产生杂音适应模型；说话人适应参数计算装置，对上述杂音适应处置所产生的前述杂音适应模型借由上述说话人适应处理时所说话的语音实行说话人适应计算，从而计算出将上述杂音适应模型变换为杂音叠加说话人适应模型用的说话人适应参数；语音模型更新装置，在对上述记忆装置的初期语音模型借由上述说话人适应处理参数实施说话人适应处理以产生说话人适应模型，而将该说话人适应模型取代前述初期语音模型从而更新记忆于上述记忆装置。

25 所述语音识别装置，其还具备有在语音识别时实行语音识别处理的识别处理装置，并且上述杂音适应装置借由上述语音识别时的非说话期间的背景杂音对更新记忆于上述记忆装置的上述说话人适应模型实施杂音适应处理，从而产生经杂音适应处理的说话人适应模型，并且将杂音适应处理过的说话人适应模型作为识别语音用的声音模型供给至上述语音识别装置。

本发明又提供一种语音识别装置，具备有：记忆装置，具有初期语音模型；杂音适应装置，借由语音识别时的非说话期间的背景杂音对上

述记忆装置的初期语音模型实施杂音适应处理而生成杂音适应模型；识别处理装置，对照上述语音识别时的说话期间所说话的应该语音识别的语音与上述杂音适应装置所生成的上述杂音适应模型而实行语音识别；说话人适应参数计算装置，对前述杂音适应装置所生成的上述杂音适应模型借由上述应该语音识别的语音实行说话人适应计算，而算出将上述杂音适应模型变换成杂音叠加说话人适应模型用的说话人适应参数；语音模型更新装置，对上述记忆装置的初期语音模型借由上述说话人适应参数实施说话人适应处理以产生说话人适应模型，并将该说话人适应模型取代上述初期语音模型而更新记忆于上述记忆装置中。

10 所述的语音识别装置，其中前述说话人适应参数计算装置与语音模型更新装置在前述识别处理装置的识别结果的可靠度较高时，生成前述说话人适应模型而取代上述初期语音模型从而更新记忆于上述记忆装置中。

本发明还提供一种语音识别方法，包括：杂音适应处理步骤；对记忆于记忆装置的初期语音模型借由说话人适应处理时的背景杂音实施杂音适应处理而产生杂音适应模型；说话人适应参数计算处理步骤，对上述杂音适应处理步骤所产生的上述杂音适应模型借由上述说话人适应处理时所说话的语音实行说话人适应计算，而计算出将上述杂音适应模型变换成杂音叠加说话人适应模型用的说话人适应参数；语音模型更新处理步骤，对上述记忆装置的初期语音模型借由上述说话人适应参数实施说话人适应处理而产生说话人适应模型，而以该说话人适应模型取代上述初期语音模型从而更新记忆于上述记忆装置。

25 所述的语音识别方法，其在上述杂音适应处理步骤中还具备有语音识别处理步骤，借由在语音识别时的非说话期间的背景杂音，对更新记忆于上述记忆装置的上述说话人适应模型实施杂音适应处理，从而生成实施过杂音适应处理的说话人适应模型，而借由对照该实施过杂音适应处理的说话人适应模型与上述语音识别时的说话期间的应该语音识别的语音来实行语音识别。

本发明又提供一种语音识别方法，包括：杂音适应处理步骤，借由

语音识别时的非说话期间的背景杂音对记忆于记忆装置的初期语音模型实施杂音适应处理而产生杂音适应模型；识别处理步骤，对照上述语音识别时的说话期间所说话的应该语音识别的语音与上述杂音适应处理步骤所生成的上述杂音适应模型而实行语音识别；说话人适应参数计算处理步骤，对上述杂音适应处理步骤所生成的上述杂音适应模型借由上述应该语音识别的语音实行说话人适应计算，从而算出将上述杂音适应模型变换成杂音叠加说话人适应模型用的说话人适应参数；语音模型更新处理步骤，对上述记忆装置的初期语音模型以上述说话人适应参数实施说话人适应处理而产生说话人适应模型，而将该说话人适应模型取代上述初期语音模型以更新记忆于上述记忆装置。

所述的语音识别方法，其中上述说话人适应参数计算处理步骤与语音模型更新处理步骤在上述识别处理步骤的识别结果的可靠度较高的情况下产生上述说话人适应模型，从而取代前述初期语音模型而更新记忆于上述记忆装置中。

所述的语音识别装置与语音识别方法，其中，在说话人适应处理时，对初期语音模型实施杂音适应处理，而对以该杂音适应处理所得的杂音适应模型实行说话人适应计算，而计算出将杂音适应模型变换成杂音叠加适应模型用的说话人适应参数，而对初期语音模型以前述说话人适应参数实施说话人适应处理而生成在语音识别时所利用的说话人适应模型，而将初期语音模型以该说话人适应模型加以更新。

借此，可减低说话人适应处理时的背景杂音的不良影响，从而可实现说话人适应处理的本来目的，产生对说话人的个人特征的适应效果较高的说话人适应模型。

所述的语音识别装置与语音识别方法，在实行说话人适应处理后的语音识别时，借由其语音识别时的非说话期间的背景杂音对被更新记忆的说话人适应模型实施杂音适应处理。借此，产生经杂音适应处理的说话人适应模型。然后，借由对照实施杂音适应处理的说话人适应模型与语音识别时的说话期间的应该语音识别的语音，实行语音识别。

所述的语音识别装置与语音识别方法，可在语音识别时同时实行说

话人自适应。

也就是，借由语音识别时的非说话期间的背景杂音对初期语音模型实施杂音适应处理而生成杂音适应模型，而借由对照其语音识别时的说话期间的说话的应该语音识别的语音与杂音适应模型可实行语音识别。

5 还有，对杂音适应模型借由应该语音识别的语音实行说话人适应计算，可算出将杂音适应模型变换成杂音叠加说话人适应模型用的说话人适应参数。还有，借由对初期语音模型以说话人适应参数实施说话人适应处理，可产生说话人适应模型，从而以生成的说话人适应模型取代更新初期语音模型。

10 其结果是，随着对大量的语音实行语音识别，可将初期语音模型更新成对说话人的个人特征的适应程度较高的说话人自适应模型，可实现语音识别性能的提高。

所述语音识别装置与语音识别方法，其在语音识别结果的可靠度较高的情况下，借由产生说话人适应模型而更新初期语音模型，可对应说话环境的状态等实现适当的说话人适应处理。

15 还有，所谓初期语音模型，指实施说话人适应处理前记忆于前述记忆装置的语音模型。在本发明中，利用由说话人适应处理所产生的说话人适应模型将记忆于记忆装置的初期语音模型予以更新，然后，此经更新的说话人适应语音模型被作为初期语音模型处理。也就是，在记忆装置中最初记忆初期语音模型，而在以说话人适应模型更新后，将更新过的说话人适应模型视为初期语音模型加以处置而重复其操作。

附图说明

下面结合附图对本发明的具体实施方式作进一步详细的描述。

25 附图中，

图 1 为第一实施方式的语音识别装置的构成的示意图；

图 2 为正确的判断识别结果为正确或错误的原理的说明图；

图 3 为第一实施方式的语音识别装置的操作的流程图；

图 4 为第二实施方式的语音识别装置的构成的示意图；

图 5 为第三实施方式的语音识别装置的构成的示意图；
图 6 为第三实施方式的语音识别装置的操作的流程图；
图 7 为第四实施方式的语音识别装置的构成的示意图；
图 8 为第四实施方式的语音识别装置的操作的流程图。

5

具体实施方式

以下参照图示说明本发明的较佳实施方式。

(第一实施方式)

现在参照图 1 至图 3 说明本发明的第一实施方式。其中，图 1 为本
10 实施方式的语音识别装置的构成图。

在图 1 中，本语音识别装置包括有：使用 HMM 模型 (Hidden Markov Model) 实行语音识别，将作为收音装置的麦克风 (省略图示) 所输出的语音输入讯号 $v(t)$ 变换成倒频谱 (cepstrum) 范围的特征矢量系列 $V(n)$ 来输出的声音分析部 1，语音识别处理部 2，单词模型生成部 3，单词词典 4，
15 声音模型记忆部 5，正误判断部 6 及说话人适应处理部 7。

在声音模型记忆部 5 中记忆着由非特定人的语音所产生的音素等次单词单位的声音模型 (标准的音素 HMM 模型)。

还有，由后述可知，在声音模型记忆部 5 中预先将来自非特定人的语音的声音模型 (标准的音素 HMM 模型) 当作初期声音模型而加以记忆，
20 其后，在实施说话人适应处理后，其初期的声音模型被更新成经说话人适应处理的声音模型。接着，在继续说话人适应处理或者重复实行时，记忆于声音模型记忆部 5 的说话人适应声音模型被实行说话人适应处理，再次对该经说话人适应处理的声音模型实施以说话人适应处理，而重复加以处理，在声音模型记忆部 5 记忆着重复更新的说话人适应声音模型。
25 这样，借由持续更新说话人适应声音模型可实现对说话人语音的识别性能的提高。

在单词词典 4 中预先记忆着大量的单词与文句的词典资料 (text data)。

单词模型生成部 3 依照记忆于单词词典 4 中的各单词或文句的资料

(以下将这些资料统称为“单词资料”)WRD 将记忆在声音模型记忆部 5 的各音素的声音模型 HMMsb 组合,借此产生对应于单词资料 WRD 的音响模型(以下称为“单词模型”)HMMw。

声音识别处理部 2 将单词模型生成部 3 所生成的单词模型 HMMw 与自
5 声音分析部 1 所供给的语音的特征矢量 $V(n)$ 作实时(real time)对照。然后,对照全部的单词模型 HMMw 与特征矢量 $V(n)$ 的结果输出可得到最大优度揭示单词模型 HMMw 的识别结果 RCG。

还有,语音识别处理部 2 除输出识别果 RCG 外,将所述的最大优度的值作为第一记录 FSCR,而将次大的优度值(相对于最大优度为第二大的
10 优度值)作为第二记录 SSCR 输出。

以具体例子作说明,例如说话人讲出“东京(Tokyo)”时,语音识别处理部 2 对照其特征矢量系列 $V(n)$ 与全部的单词模型 HMMw 的结果,“东京(Tokyo)”的单词模型的优度成为最大值,而其它的单词“京都(Kyoto)”的单词模型的优度如果为第二大的优度值时,则将其最大优度的值作为
15 第一记录 FSCR 而将第二大的优度值作为第二记录 SSCR 输出。还有,在前述例子中,“东京(Tokyo)”的语音与相当于第一记录 FSCR 的“东京(Tokyo)”的单词模型如果一致,则识别结果 RCG 为相当于正确的情况。

另一方面,识别结果 RCG 为误识别的情况,例如说话人讲出“东京(Tokyo)”时,语音识别处理部 2 识别语音的结果例如单词“京都(Kyoto)”
20 的单词模型的优度为最大值,而“东京(Tokyo)”的单词模型的优度为第二大值时,其最大的优度值被作为第一记录 FSCR 而第二大的优度值被作为第二记录 SSCR 输出。这样,因“东京(Tokyo)”的语音与相当于第一记录 FSCR 的“京都(Kyoto)”的单词模型并不一致,因此识别结果 RCG 变得不正确。

正误判断部 6 将语音识别处理部 2 对某一语音作语音识别时的输出的第一记录 FSCR 与第二记录 SSCR 应用于下式(1)所表示的记录评价函数,然后将其所得的记录评价 $G(L)$ 与一定的基值 THD 作比较。

$$G(L)=FSCR \times (FSCR-SSCR) \quad \dots(1)$$

其中,上述式(1)的右边中所示的变量 FSCR 为第一记录 FSCR 的值(最

大优度值), 而变量 SSCR 为第二记录 FSCR 的值(第二的优度值), 记录评价价值 $G(L)$ 为与借由右边的评价计算所得到的优度 L 有关的变量。

还有, 当记录评价价值 $G(L)$ 比基值 THD 大时(也就是 $G(L) > \text{THD}$ 时), 则将识别结果 RCG 判断为正确, 而当记录评价价值 $G(L)$ 小于基值 THD 时(也就是 $G(L) \leq \text{THD}$ 时), 则识别结果 RCG 判断为误, 从而分别输出判断结果 RSLT。

在此, 详细叙述上述式(1)的记录评价函数与基值 THD 以及正误判断部 6 的判断原理。

上述式(1)的记录评价函数与基值 THD 是借由下述的统计方法预先实验决定出。

首先, 令任意的说话人作一定数量 N 的单词以及文句的说话, 而借由单词辞典 4 与单词模型生成部 3 与语音识别处理部 2 实行语音识别, 而实验性的测量总数 N 的各单词及文句所输出的第一记录 FSCR 以及第二记录 SSCR。

然后, 分类成语音与识别结果 RCG 为一致时的单词及文句(也就是被正确识别时的单词或文句)与语音与识别结果 RCG 为不一致时的单词及文句(也就是被误识别时的单词及文句)。

例如作为实验对象其所说话的单词及文句的总数 N 为 500 个, 其中被正确识别的单词及文句的个数为 400 个, 而被误识别的单词及文句的个数 Y 为 100 个, 则将 500 个单词及文句分别分类成个数分别为 X 及 Y 的两类。

然后, 如图 2(a)所示, 利用被正确地识别的单词或文句(在已述的例子中为 $X=400$ 个单词及文句), 作成第一记录 FSCR 的值(最大优度值)相对应的单词及文句的个数的分布的曲线 $P(\text{FSCR})$, 并作成第二记录 SSCR 的值(第二优度值)相对应的单词及文句的个数的分布的曲线 $Q(\text{SSCR})$ 。

也就是, 曲线 $P(\text{FSCR})$ 与 $Q(\text{SSCR})$ 以被正确的语音识别的 400 个单词及文句为对象所作成。还有, 语音识别处理部 2 所输出的第一记录 FSCR 的值(最大优度值)与第二记录 SSCR 的值(第二的优度值)对应于语音识别环境等作各种变化, 因此将 400 个单词及文句相对应于第一记录 FSCR 的

各个值作分配而作成图中以实线所示的曲线 P(FSCR)，同样的借由将 400 个单词及文句相对应于第二记录 SSCR 的各个值作分配而作成图中以虚线所示的曲线 Q(SSCR)。

还有，如图 2(b) 所示，同样利用被误识别的单词及文句(在已述的例子中为 $X=100$ 个单词及文句)，作成揭示相对应于第一记录 FSCR 的值的单词及文句的个数的分布的曲线 P(FSCR)”与揭示相对应于第二记录 SSCR 的值的单词及文句的分布的曲线 Q(SSCR)”。

也就是，图 2(b) 所示的曲线 P(FSCR)”与 Q(SSCR)”以误识别的 100 个单词及文句为对象作成。还有，语音识别处理部 2 在误识别的情况下，
10 前述第一记录 FSCR 的值(最大优度值)与第二记录 SSCR 的值(第二优度值)相对应于语音识别环境等作各种变化，因此借由将 100 个单词及文句相对应于第一记录 FSCR 的各个值作分配，作成图 2(b) 中的实线所示的曲线 P(FSCR)”，而借由将 100 个单词与文句相对应于第二记录 SSCR 的各个值作分配，作成图 2(b) 中的虚线所示的曲线 Q(SSCR)”。

15 这样，在作成曲线后，图 2(a) 中的曲线 P(FSCR) 与 Q(SSCR) 偏向分布在各个分离的优度值范围，因此在识别结果 RCG 为正确的情况下，第一记录 FSCR 的统计性特征与第二记录 SSCR 的统计性特征表现出很大的差异。

还有，图 2(b) 中的曲线 P(FSCR)”与 Q(SSCR)”分布在大约相同的优
20 度值的范围，因此，在识别结果 RCG 为错误的情况下，第一记录 FSCR 的统计性特征与第二记录 SSCR 的统计性特征大约相同。

这样，在识别结果 RCG 为正确的情况下及在识别结果为错误的情况下，曲线 P(FSCR) 与 Q(SSCR) 以及曲线 P(FSCR)”与 Q(SSCR)”的关系具有特有的统计特征，作为表现该统计特征的函数被确定为上述式(1)的记录
25 评价函数。

依照上述式(1)的记录评价函数，在识别结果 RCG 为正确的情况下，如图 2(a) 所示，第一记录 FSCR 比第二记录 SSCR 偏向于较大的优度值侧，上述式(1)的右边中的差值 (FSCR-SSCR) 很大，而借由对此差值 (FSCR-SSCR) 乘以第一记录 FSCR 可得到将其差值 (FSCR-SSCR) 的特征更显

著化的记录评价值。

因此，上述式(1)的记录评价函数借由用于语音识别正确实行时的第一记录 FSCR 与第二记录 SSCR，可适当地反映出正确地实行语音识别时的统计性特征，而且还可作为判断正确实行语音识别时所输出的识别结果 RCG 的基础。

另一方面，在实行误识别的情况下，如图 2(b)所示，因第一记录 FSCR 与第二记录 SSCR 的优度值大约产生于相同范围内，因此上述式(1)的右边中的差值 (FSCR-SSCR) 很小，再对此差值 (FSCR-SSCR) 乘以第一记录 FSCR 可将其差值 (FSCR-SSCR) 的特征更加显著化，从而可得到明确的记录评价值。

因此，上述式(1)的记录评价函数借由用于实行误识别时的第一记录 FSCR 与第二记录 SSCR 可适当的反映实行误识别时的统计特征，从而可作为实行误识别时判断输出的识别结果 RCG 为错误的基础。

其次，基值 THD 以自上述式(1)的记录评价函数所求得的记录评价值 $G(L)$ 为对象，再识别结果 RCG 为正确时判断为正确，而再识别结果为误时判断为错误，从而可成为判断的基准。

也就是，想要将全部的识别结果 RCG 的正确解答及不正确解答正确的予以判断一般很困难。在说话人适应处理时，若将不正确解答误判断为正确解答，则如前述，声音模型将会被错误地实行适应处理，将导致识别性能的降低。相反的，将正确解答误判断为不正确解答时，因不实行说话人适应处理，故识别性能不能改善但也不降低。因此，基值 THD 为充份确保将不正确解答判断为不正确的性能，根据下述的原理加以选择。

首先，将正确实行语音识别时的各单词及各文句所取得的各个第一记录 FSCR 与第二记录 SSCR 以上述式(1)的记录评价函数加以运算，而算出各单词及各文句的记录评价值 $G(L)$ 。然后，求出对应于所算出的各记录评价值 $G(L)$ 的单词及文句的个数的分布的曲线 $R(G(L))$ 。

同样的，将实行误识别时的各个单词及文句所得的第一记录 FSCR 与第二记录 SSCR 以上述式(1)的记录评价函数加以运算，而算出各单词及

各文句的记录评价值 $G(L)$ 。然后, 求出对应于所算出的各记录评价值 $G(L)$ 的单词及文句的个数的分布的曲线 $T(G(L))$ 。

这样, 在求出各曲线 $R(G(L))$ 与 $T(G(L))$ 后, 曲线 $R(G(L))$ 为图 2(c) 中的实线所示的分布状态, 而曲线 $T(G(L))$ 为图 2(c) 中的虚线所示的分布状态。

也就是, 曲线 $R(G(L))$ 表示正确识别某单词或某文句时的特征, 而曲线 $T(G(L))$ 表示误识别某单词或某文句时的特征。

还有, 以曲线 $R(G(L))$ 的个数为 0 时的记录评价值 $G(L)$ 为界限, 比其记录评价值 $G(L)$ 的值大的范围 WR 在识别结果 RCG 为正确时成为可判断其为正确的范围, 而以曲线 $T(G(L))$ 的个数为 0 时的记录评价值 $G(L)$ 为界限, 比其记录评价值 $G(L)$ 的值小的范围 WT 成为在识别结果 RCG 为误时可判断其为错误的范围。

为此, 基值 THD 被决定为比曲线 $T(G(L))$ 的个数为 0 时的记录评价值 $G(L)$ 稍大一些的值。

还有, 正误判断部 6 在每次输出各单词及各文字的识别结果 RCG 时比较基值 THD 与由上述式(1)的记录评价函数所取得的记录评价值 $G(L)$, 来判断基值 THD 与记录评价值 $G(L)$ 是否为既定的关系, 并输出其判断结果 $RSLT$ 。也就是, 在记录评价值 $G(L)$ 比基值 THD 大时 ($G(L) > THD$ 时), 则判断识别结果 RCG 为正确, 而在记录评价值 $G(L)$ 比基值 THD 小时 ($G(L) \leq THD$ 的时), 则判断识别结果 RCG 为错误, 而分别输出判断结果 $RSLT$ 。

说话人自适应处理部 7 输入识别结果 RCG 与判断结果 $RSLT$, 然后对应于判断结果 $RSLT$ 实行说话人适应处理。

也就是, 在输入表示正确的语音识别的判断结果 $RSLT$ 时, 借由在那时的说话人的语音的特征矢量 $V(n)$ 实行说话人适应处理, 另一方面, 在输入的判断结果 $RSLT$ 为误判断时不实行说话人适应处理。

还有, 上述的说话人适应处理是对说话人适应处理前的音素模型 $HMMsb$ 的全部或者一部份加以实行。

也就是, 输入表示正确的语音识别的判断结果 $RSLT$ 时, 语音的特征矢量系列 $V(n)$ 的说话内容被视为识别结果 RCG , 而借由以说话内容为已

知条件的 MLLR 或 MAP 算法等说话人适应处理方法，对说话人适应处理前的音素模型 HMMsb 实施说话人适应处理而取得说话人适应处理后的因素模型 HMMsb”。然后，将此 HMMsb” 供给至声音模型记忆部 5，从而置换说话人适应处理前的声音模型 HMMsb 而更新记忆。

- 5 然后，不必说，说话人适应处理被持续性地或者重复性地实行。因此，更新记忆于声音模型记忆部 5 中的说话人适应处理后的音素模型成为接下来的说话人适应处理过程的说话人适应处理前的音素模型，而该说话人适应处理所得到的音素模型被更新记忆于声音模型记忆部 5 后，在下回实行说话人适应处理的时候对被更新的音素模型实施说话人适应
- 10 处理，然后用这种方式重复进行。

接着参照图 3 所示的流程图，说明具有前述构成的本语音识别装置的操作。

在本图中，开始说话人适应处理后，在步骤 S100 开始语音输入的处理。

- 15 接着，在步骤 S102，单词词典 4 与单词模型生成部 3 及语音识别处理部 2 对照语音的特征矢量 $V(n)$ 与单词模型 HMMw 来实行语音识别，识别结果 RCG 与第一记录 FSCR 及第二记录 SSCR 被输出。

接着在步骤 104，正误判断部 6 将第一记录 FSCR 与第二记录 SSCR 应用于上述式(1)的记录评价函数中，从而算出记录评价值 $G(L)$ 。

- 20 接着在步骤 106，正误判断部 6 比较记录评价值 $G(L)$ 与基值 THD，而在 $G(L) > THD$ 时(“YES”时)，判断识别结果 RCG 为正确而转移至步骤 S108 的处理，而在 $G(L) \leq THD$ 时(“NO”时)判断识别结果 RCG 为误，而不实施说话人适应处理就转移至结束。

- 处理转移至步骤 S108 后，说话人适应处理部 7 借由特征矢量 $V(n)$
- 25 对声音模型记忆部 5 中的音素模型 HMMsb 作说话人适应处理，然后在步骤 S110 将说话人适应处理后的音素模型 HMMsb” 更新记忆，之后终止处理。

还有，在图 3 中为方便的原因，只揭示了针对说话的一个单词或一句文句的语音实施说话人适应处理的情形，在由许多单词或文句所构成

的文章成为对象而实施说话人适应处理的情况下，只要重复图 3 的处理过程就可以。

然后，顺序处理被说话的大量的单词以及文句，而对于误识别的单词及文句的音素模型 HMMsb 在步骤 106 中被判断为“NO”，因此跳过说话人适应处理(步骤 108 及步骤 110)，而对正确识别的单词及文句在步骤 106 中被判断为“YES”，而借由实施说话人适应处理(步骤 108 及步骤 110)实行适当的说话人适应处理。

这样，依照本实施方式的语音识别装置，借由将两个第一记录 FSCR 与两个第二记录 SSCR 应用于记录评价函数中，可得到记录评价值 $G(L)$ ，而将该值与既定的基值 THD 作比较时可确实并且迅速的判断识别结果 RCG 为正确或错误。也就是，可大量削减判断语音识别的结果 RCG 为正确或错误的处理量，同时可获得高判断准确度。

因此，可在不使说话人感到压力的情形下迅速并且适当的实行说话人适应处理。同时，因为可大幅减少错误的说话人适应处理，因此可防止说话人适应处理后的语音识别精度变差等问题的发生。

还有，成为说话人的使用者在每次使用本语音识别装置时，可对声音模型记忆部 5 中的音素模型 HMMsb 逐渐的实施适当的说话人适应处理，所以可随着使用次数的增加将语音识别率予以提高。

还有，在本实施方式中如前述式(1)所示根据第一记录 FSCR 与第二记录 SSCR 计算记录评价值 $G(L)$ ，其先求取第一记录 FSCR 与第二记录 SSCR 的差值，然后还对其差值乘以第一记录 FSCR。但是，本发明并不只限定于此，作为变化的例子也可将第一记录 FSCR 与第二记录 SSCR 的差值作为记录值 $G(L)$ 。

25 (第二实施方式)

接着参照图 4 说明本发明的第二实施方式。其中，在图 4 中与图 1 相同或相当的部份标注以相同符号。

本语音识别装置与图 1 所示的实施方式的语音识别装置作对比时，在本语音识别装置中还具有成为区别的误判断对应部 8 与显示部 9。

误判断对应部 8 输入来自正误判断部 6 的判断结果 RSLT，而在识别结果 RCG 的判断结果 RSLT 为正确时输出该识别结果 RCG，而在识别结果 RCG 的判断结果 RSLT 为误识别时则禁止该识别结果 RCG 的输出。借此，只在正确的实行语音识别的情况下才输出识别结果 RCG。

- 5 还有，在识别结果 RCG 的判断结果 RSLT 为误识别时，借由指示于由液晶显示器等所形成的显示部 9 上，将表示误识别与应实行再次说话含义的文字等所构成的警告予以显示出。

这样，依照本发明的语音识别装置，其借由利用正误判断部 6 的判断结果 RSLT 可对使用者提示是否已实行了适当的说话人适应处理，或者
10 再次实行说话等信息，可达到提高使用者的易用性的目的。

如以上所说明，依照本发明的语音识别装置以及语音识别方法，其在说话人适应处理时借由对照声音模型与语音的特征矢量，求出表示得到最大优度的声音模型的识别结果与表示最大优度值的第一记录与表示其次优度值的第二记录，而借由比较由前述第一记录与第二记录所产生的
15 的评价值与预先设定好的基值，来实行识别结果的正误判断，因此可实现高准确度并且较少处理量的正误判断。

接下来，参照图示说明关于本发明的其它较佳实施方式。

(第三实施方式)

- 20 参照图 5 及图 6 说明本发明的第三实施方式。图 5 为本实施方式的语音识别装置的构成图。

接下来，作为一较佳的实施方式，现说明在初期设定就实行说话人适应处理，而在识别时使用初期设定的语音模型加以实施的实施方式。

- 25 图 5 中，本语音识别装置使用 HMM 模型 (Hidden Markov Model) 实行语音识别，具备有：初期语音模型记忆部 11，记忆有预先在无杂音的环境下收录成的语音数据库，并且使用该数据库作学习的成为非特定人语音模型的初期语音模型 M_c ；说话人自适应模型记忆部 12，用来更新记忆后述的说话人适应计算处理所生成的说话人适应模型 M_c ；杂音适应部 13；说话人适应部，具备有说话人适应参数计算部 14 与语音模型更新部

15。

还具备有：声音分析部 16，将麦克风 17 所收音的收音讯号 $v(t)$ 转换成在各既定的帧 (frame) 期间的倒频谱系数矢量，从而生成倒频谱范围的特征矢量系列 $V(n)$ ；切换开关 18 以及一识别处理部 19。

5 还有，同图中以带箭头的虚线与带箭头的实线以模式化的方式揭示讯号的通过路径，而带箭头的虚线表示仅在说话人适应处理时通过的讯号的流程，另一方面，带箭头的实线表示仅在语音识别时所通过的讯号的流程，或者语音识别时以及说话人适应处理时的讯号的流程。

在此，声音分析部 16 借由将麦克风 17 作收音所产生的时间轴上的
10 收音讯号 $v(t)$ 在每个既定的帧时间转换成倒频谱系数矢量，而产生并输出倒频谱范围的特征矢量系列 $V(n)$ 。其中，符号 $V(n)$ 中的变量 n 表示帧号码。

切换开关 18 在后述的说话人适应处理之际，在说话人尚未开始说话的非说话期间内切换至杂音适应部 13 一侧，从而将语音分析部 16 所生成
15 的特征矢量系列 $V(n)$ 送出至杂音适应部 13。

这部份看作说话人或者装置在处理开始的指示被执行后的数十微秒乃至数秒之间语音尚未发出的区间，将切换开关 18 切换至杂音适应部 13 一侧，其后将切换开关 18 切换控制至说话人自适应计算部 14 或者识别处理部 19 一侧而加以实现。

20 或者，将来自麦克风 17 的收音讯号 $v(t)$ 以既定的监视控制装置 (省略图示) 作逐次监视，在由收音讯号 $v(t)$ 中判断无说话人的语音时将切换开关 18 切换至杂音适应部 13 一侧，而在判断出已有说话人的语音时则切换控制至说话人适应计算部 14 或者识别处理部 19 一侧。重点在于对应其是否为收音讯号 $v(t)$ 中含有说话人的语音的区间而将切换开关 18 如
25 上述般切换。

还有，在说话人适应处理之际，当说话人开始说话时，在其说话期间内切换至说话人适应参数计算部 14 一侧，而将语音分析部 16 所生成的特征矢量系列 $V(n)$ 送出至说话人适应参数计算部 14。

接着，在说话人适应处理终了后，在语音识别处理开始的情况下，

切换开关 18 也实行同样的操作，然而如前述般，讯号并不通过带箭头的虚线所示的路径而通过带箭头的实线所通过的路径。

在说话人开始说话前仅收录背景杂音期间，切换开关 18 切换至杂音适应部 13 一侧，而将背景杂音的特征矢量 $N(n)$ 送出至杂音适应部 3。

5 在到达麦克风 17 收录语音的说话期间时，切换开关 18 在说话人适应处理时切换至说话人自适应参数计算部 14 一侧，而在识别处理时则切换至识别处理部 19 一侧，将说话期间的特征矢量系列 $V(n)$ 送出至说话人自适应参数计算部 14 或者识别处理部 19。

10 初期语音模型记忆部 11 为由只读半导体内存 (ROM) 或者设为易装卸的智能型媒体或者小型闪存等所形成的数据库，其记忆借由预先学习标准的说话人语音所产生的音素等次单词单位的初期语音模型 M_c 。

说话人适应模型记忆部 12 由可反复记忆的半导体内存等所形成，在后述的说话人适应处理之际，首先写入记存于初期语音模型记忆部 1 中的初期语音模型 M_c 。

15 然后则如后述，借由说话人适应参数计算部 14 与语音模型更新部 5 将 HMM 模型作说话人适应处理，而在以说话人适应模型 M_c'' 更新后，将初期语音模型 M_c 置换 (更新) 为说话人适应模型 M_c'' 。

20 杂音适应部 13 在说话人适应处理之际对记忆于说话人适应模型记忆部 12 的音素等次单词单位的全部的初期语音模型 M_c 实施杂音适应处理，借以生成对应于全部的初期语音模型 M_c 的杂音适应模型 M_c' ，并经由同图中的带箭头的虚线所示的路径送出至说话人适应参数计算部 14。

25 还有，杂音适应部 13 在语音识别时借由上述说话人适应处理对更新记忆于说话人适应模型记忆部 12 中的语音模型 (也就是说话人适应模型 M_c'') 实施杂音适应处理，而将该经杂音适应处理的说话人适应模型 M_{reg} 经由同图中带箭头的实线所示的路径送出至识别处理部 19。

也就是，在前者的说话人适应处理之际，在说话人适应处理时在说话人尚未说话的非说话期间，在将其说话环境中所产生的背景杂音以麦克风 17 收音之后，声音分析部 16 由其收音讯号 $v(t)$ 产生各既定帧期间的特征矢量系列 $V(n)$ ，并且切换开关 18 切换至杂音适应部 13 一侧，将

前述特征矢量系列 $V(n)$ 当作背景杂音的特征矢量系列 $N(n)$ 送出至杂音适应部 13。

还有，杂音适应部 13 使用其背景杂音特征矢量系列 $N(n)$ ，自所有的初期语音模型 M_c 中借由 HMM 模型合成方法或者雅可比(Jaboci)适应方法等杂音适应处理产生杂音适应模型 M_c' ，而送出至说话人适应参数计算部 14。

还有，在后者的语音识别处理之际，在语音识别时，在说话人尚未说话的非说话期间，将该说话环境所产生的背景杂音以麦克风 17 加以收音，声音分析部 16 由该收音讯号 $v(t)$ 产生各既定帧期间的特征矢量系列 $V(n)$ ，而切换开关 18 借由切换至杂音适应部 13 一侧，将其特征矢量系列 $V(n)$ 作为背景杂音的特征矢量系列 $N(n)$ 送出至杂音适应部 13。

接着，杂音适应部 13 使用前述背景杂音特征矢量系列 $N(n)$ 对经过更新记忆的语音模型(也就是说话人自适应模型 M_c'') 实施杂音适应处理，而将经杂音适应处理的说话人适应模型 M_{reg} 送出至识别处理部 19。

在此，现说明杂音适应部 13 在说话人适应处理时使用 HMM 模型合成法将对语音识别率具有巨大影响的说话人适应模型 M_c'' 的各分布的平均矢量 μ_c 作杂音适应处理的情形。

首先，杂音适应部 13 自背景杂音的特征矢量系列 $N(n)$ 求取背景杂音模型 N_b 。

在此，为说明的方便，假设背景杂音为定常状态，而背景杂音模型 N_b 为一状态一混合模型，而在将背景杂音模型 N_b 的平均矢量作为 μ_N 说明时，平均矢量 μ_N 以帧数将背景杂音的特征矢量系列 $N(n)$ 加以平均来求得。

接着，借由合成初期语音模型 M_c 的分布 m 的平均矢量 μ_{cm} 与背景杂音模型 N_b 的平均矢量 μ_N 求得下式(2)所表示的合成后的杂音适应分布 m 的平均矢量 μ_{cm}' 。

$$\mu_{cm}' = \text{IDCT}[\log[(\exp[\text{DCT}[\mu_{cm}]] + k \cdot \exp[\text{DCT}[\mu_N]])]] \quad \dots (2)$$

在此，IDCT[]为逆离散余弦变换，log[]为对数变换，exp[]为指数变换，DCT[]为离散余弦变换， k 为自 SN 比求得的混合比。

对初期语音模型 M_c 的全部分布求取前述值。借此，求出在初期语音模型 M_c 上叠加以说话人适应处理时的说话环境下的背景杂音的形态的杂音适应模型 M_c' ，并送出至说话人适应参数计算部 14。

接着，在此虽将杂音模型定为一状态一混合，而在两状态以上或两混合以上的情况下则对初期语音模型 M_c 的一分布求取大量的对应的杂音适应模型 M_c' 的分布。还有，考虑共分散行列的情况下时也可求取杂音适应模型 M_c' 。

使用 HMM 模型合成法作为杂音适应方法只是用作说明，在本发明中也可使用雅可比适应方法或其它的求取在初期语音模型 M_c 上叠加以说话时的背景杂音的状态的杂音适应模型 M_c' 用的杂音适应方法。

说话人适应参数计算部 14 在说话人适应处理时，从杂音适应部 13 输入杂音适应模型 M_c' ，从自声音分析部 16 输入，介以切换开关 18 作供给的语音的特征矢量系列 $V(n)$ ，从而生成输出具有语音特征的说话人适应参数 P 。

更具体的描述，在说话人适应处理时，在说话人开始说话时，在其说话期间，切换开关 18 被切换至说话人适应参数计算部 14 一侧，而叠加上有背景杂音的语音的特征矢量系列 $V(n)$ ，从语音分析部 16 经由切换开关 18 被供给至说话人适应参数计算部 14。

这样，叠加上有背景杂音的语音(背景杂音叠加语音)的特征矢量系列 $V(n)$ ，与和其同样经过背景杂音之杂音适应处理的杂音适应模型 M_c' 被予以供给后，说话人适应参数计算部 14 使用这些特征矢量系列 $V(n)$ 与杂音适应模型 M_c' 实行说话人适应计算处理，而产生将杂音适应模型 M_c' 作说话人适应处理用的说话人适应参数 P 。

在此，使用 MLLR(Maximum Likelihood Linear Regression)算法为例作为对说话人适应计算处理的说明，来说明对识别率具有很大影响的说话人适应模型 M_c 的各分布的平均矢量作更新的情形。

说话内容使用已知的语音的特征矢量系列 $V(n)$ 与杂音适应模型 M_c' 实行 MLLR 处理，而将杂音适应模型 M_c' 的分布 m 的平均矢量 μ_{cm}' 作为说话人适应处理用的说话人适应参数 P ，从而求出变换行列 W_m' 与

偏移矢量 bm' 。

在此，变换行列 Wm' 与偏移矢量 bm' 因为共有大量的分布，所以在若干的分布中使用相同值的变换行列 Wm' 与偏移矢量 bm' 。

还有，共有变换行列 Wm' 与偏移矢量 bm' 的分布的选择借由将全平均矢量加以分组，从而预先计算原本的杂音适应处理前的分布。

接着，在全部的分布中为共有变换行列 Wm' 与偏移矢量 bm' 的情况下，对全分布求取共通的一类的变换行列 Wm' 与偏移矢量 bm' 。

接着，在上述的杂音适应部 13 所使用的杂音模型非为一状态一混合时，杂音适应模型 Mc' 的大量分布对应于初期语音模型 Mc 的一个分布，此情况下，在对应于初期语音模型 Mc 的一分布的全部的杂音适应模型 Mc' 上，共有变换行列 Wm' 与偏移矢量 bm' 。

还有，MLLR 算法一般使用多份说话内容的语音资料实行计算，在这里是将说话人适应参数 P 的分布间的共有信息在全体说话内容中共通使用，对应于语音资料的声音模型使用对各说话内容实行杂音适应的杂音适应模型 Mc' 作计算。

这样，在使用 MLLR 算法作为说话人适应方法的情况下，在说话人适应参数计算部 14 中，说话内容使用已知的语音的特征矢量系列 $V(n)$ 作为将声音模型 Mc' 的各分布的平均矢量作更新用的说话人适应参数 P ，求取变换行列 Wm' 与偏移矢量 bm' 。

还有，如上述般，虽举例说明使用 MLLR 算出变换行列 Wm' 与偏移矢量 bm' 的情形，但是其也可适用 MAP (Maximum A Posteriori) 算法。

采用此 MPA 算法求取适应平均矢量 $\mu_{cm'}$ 用的参数 P 时，借由 MAP 算法将杂音适应模型 Mc' 的平均矢量作说话人适应处理，而借由说话人适应参数计算部 14 由其处变换出说话人适应参数 P 。

此 MAP 算法，为将说话内容借由已知的语音特征矢量系列 $V(n)$ 的各帧的特征矢量与杂音适应模型 Mc' 的各分布的对应关系借由维托毕 (Viterbi) 整合等加以算出。

之后，搜集对应于杂音适应模型 Mc' 的分布 m 的帧的特征矢量，而借由将其以帧数作平均而求出平均特征矢量 Vm^{\sim} 。

此时，将对应于分布 m 的帧的特征矢量的帧数(个数)设为 n_m ，而将分布 m 的重度系数设为 τ_m' ，而将把分布 m 的平均矢量 μ_{cm}' 作了说话人适应处理的更新平均矢量设为 $\mu_{cm}'^{\wedge}$ 时，则可依下式(3)所表示的关系算出其更新平均矢量 $\mu_{cm}'^{\wedge}$ 。

$$\mu_{cm}'^{\wedge} = \frac{\tau_m' \cdot \mu_{cm}' + n_m \cdot V_m}{\tau_m' + n_m} \quad \dots(3)$$

还有，重度系数 τ_m' 也以下式(4)所表示的关系对各说话内容作更新。

$$\tau_m'^{\wedge} = \tau_m' + n_m \quad \dots(4)$$

之后，以更新平均矢量 $\mu_{cm}'^{\wedge}$ 置换平均矢量 μ_{cm}' ，再将重度系数也以 $\tau_m'^{\wedge}$ 将 τ_m' 置换，而再每次说话时将平均矢量 μ_{cm}' 与重度系数 τ_m' 分别以更新平均矢量 $\mu_{cm}'^{\wedge}$ 与重度系数 $\tau_m'^{\wedge}$ 顺序加以更新。

在此，再将说话人适应参数 P 当作或者适应处理后的模型与说话人适应处理前的模型的差矢量考虑时，成为分布 m 的说话人自适应参数 P 的差矢量 dm' 可用下式(5)表示。

$$dm' = \mu_{cm}'^{\wedge} - \mu_{cm}' = \frac{n_m \cdot (V_m - \mu_{cm}')}{\tau_m' + n_m} \quad \dots(5)$$

依照此式(5)时，可不需算出更新平均矢量 $\mu_{cm}'^{\wedge}$ 而求出差矢量 dm' 。

接着，将差矢量 dm' 转送至后述的语音模型更新部 15，并将重度系数 τ_m' 以上述式(4)更新而收容于说话人适应参数计算部 14 中。还有，重度系数 τ_m' 的初期值可选择任意的值。

还有，上述杂音适应部 13 所使用的杂音适应模型 M_c' 不是一状态一混合时，杂音适应模型 M_c' 的大量分布对应于初期语音模型 M_c 的一个分布。

例如初期语音模型 M_c 的分布 m ，对应于杂音适应模型 M_c' 的分布

m_1, m_2, \dots, m_k 。还有, 使对应于杂音适应模型 M_c' 的分布 m_1 的由上述式(5)求出的说话人适应参数为 dm_1' , 并使重度系数为 τ_{m_1}' 时, 则可将更新初期语音模型 M_c 的分布 m 用的说话人适应参数 dm' 借由下式(6)所表示的运算处理加以求得。

$$5 \quad dm' = \frac{dm_1' + dm_2' + \dots + dm_k'}{k} \quad \dots(6)$$

还有, 借由对上述式(6)以重度系数 τ_{m_1}' 作权重的下式(7)所表示的运算将 k 个参数统合而算出说话人适应参数 dm' 也可以。

$$10 \quad dm' = \frac{\tau_{m_1}' \cdot dm_1' + \tau_{m_2}' \cdot dm_2' + \dots + \tau_{m_k}' \cdot dm_k'}{\tau_{m_1}' + \tau_{m_2}' + \dots + \tau_{m_k}'} \quad \dots(7)$$

以上说明使用 MLLR 与 MAR 算法为说话人适应方法说明说话人适应参数计算部 14 的动作。

15 其中, 作为说话人适应方法, 也可使用其它的方法。

像 MLLR 的变换行列 W_m' 与偏移矢量 b_m' 一样, 在使用借由说话人适应处理求取说话人适应参数 P 的说话人适应方法的情况下, 使用前述的说话人适应参数, 如 MAP 算法一般, 而在不直接使用说话人适应参数的情况下, 考虑对杂音适应模型 M_c' 实行说话人适应处理的说话人杂音
20 适应模型, 而借由使用成为杂音适应模型 M_c' 与杂音适应模型 M_c' 之差的说话人适应参数 P 时, 可对应于许多种说话人适应方法。

而且, 此例中虽然为平均矢量的情况下, 也可应用于适应共分散行列的情况下。

25 还有, 在多种的说话人自适应方法中, 需要知到说话内容(所说话的单词或文句为何内容)。此情况下, 在实行语音识别处理前仅实行说话人适应处理, 在这时, 应说话内容被预先决定, 而将既定的内容对说话人提示, 而根据提示的说话内容而加以处理。

在说话人适应处理中, 除适应说话人的个人特征外, 还需对说话环境作适应。

使用在无背景杂音的环境下的说话内容，而在对应使用于背景杂音的环境下所收录的语音数据库作学习的非特定人模型的初期语音模型 M_c 作说话人适应处理的情况下，因为可不受背景杂音的影响，因此只实行对说话人的个人特征的适应处理。

- 5 当使用于说话人适应处理的说话内容为在有背景杂音的环境下实行，而将其使用于上述的初期语音模型 M_c 的说话人适应处理时，可同时实行对说话人的个人特征的适应处理与对适应说话时的背景杂音的适应处理。

因此，一般在使用说话人适应处理后的说话人适应模型作语音识别
10 时，若语音识别时的说话环境与适应说话时为具有相同的杂音环境时，则可得到较高的识别率，而当实行识别的说话环境与适应说话时为不同的情况下则有可能无法获得较高的识别率。

在本发明中为解决前述问题，在实行说话人适应处理前，如上述般借由以杂音适应部 13 作杂音适应，可生成将上述初期语音模型 M_c 适应
15 于与适应处理时的语音为相同的背景杂音环境中的杂音适应模型 M_c' ，其后，在说话人适应参数计算部 14 中使用该杂音适应模型 M_c' ，实行说话人适应处理而算出说话人适应参数 P 。

还有，杂音适应模型 M_c' 在实行说话人适应处理前因已适应于说话环境为相同的背景杂音，因此以说话人适应处理所求取的说话人适应参
20 数 P ，可减轻背景杂音适应项的影响，从而可包含更多的本来目的的对说话人的个人特征的适应项。

使用此说话人适应参数 P 并且利用后述的语音模型更新部 15 将初期语音模型作更新时，可生成适应说话时的背景杂音的影响较少的说话人适应模型 M_c'' 。

- 25 语音模型更新部 15 将记忆于说话人适应模型记忆部 12 的初期语音模型 M_c ，使用说话人适应参数计算部 14 的输出的说话人适应参数 P ，变换成说话人适应模型 M_c'' 。

现在说明采用上述的 MLLR 与 MAP 算法，更新初期语音模型 M_c 的分布 m 的平均矢量 μ_{cm} 的情况下的语音模型更新部 15 的功能。

如上述般，在说话人适应参数计算部 14 中使用 MLLR 作为说话人适应处理，并且使用变换行列 Wm' 与偏移矢量 bm' 作为说话人适应参数 P 的情况下，说话人适应更新后的说话人适应模型 Mc'' 的分布 m 的平均矢量 $\mu_{cm''}$ 由下式(8)的关系求得。

$$5 \quad \mu_{cm''} = Wm' \cdot \mu_{cm} + bm' \quad \dots (8)$$

又，在说话人适应计算部 14 使用 MAP 算法作适应处理而使用差矢量 dm' 作为说话人适应参数 P 的情况下，其平均矢量 $\mu_{cm''}$ 可由下式(9)的关系求得。

$$\mu_{cm''} = \mu_{cm} + dm' \quad \dots (9)$$

10 不论是何种情况下，平均矢量 $\mu_{cm''}$ 如上述般其成为适应说话时的背景杂音的影响较少，并且可适应说话人的个人特征的平均矢量。

其后，如上述般，语音模型更新部 15 将记忆于说话人适应模型记忆部 12 中的语音模型 Mc 使用说话人适应参数生成部 14 所输出的说话人适应参数 P 加以更新，而将更新后的说话人适应模型 Mc'' 更新记忆于说话人适应模型记忆部 12。也就是，在语音识别时应使用说话人适应模型 Mc'' 15 作为语音模型 Mc 而加以更新记忆。

识别处理部 19 设置用以实行语音识别处理。也就是，在语音识别之际，杂音适应部 13 对被更新记忆于说话人适应模型记忆部 12 的说话人适应模型 Mc (也就是说话人适应模型 Mc'') 以在识别说话杂音环境下的背景杂音的特征矢量系列 $N(n)$ 实施杂音适应，从而产生经杂音适应的说话人适应模型 $Mreg$ ，而将该说话人适应模型 $Mreg$ 供给至识别处理部 19。 20

接着，识别处理部 19 对照由经过杂音适应的说话人适应模型 $Mreg$ 所构成的系列与由声音分析部 16 一侧所供给的应该识别的语音的特征矢量系列 $V(n)$ 与其识别候补单词及文句等模型，从而将由能取得最大优度的说话人自适应模型 $Mreg$ 所构成的系列当作识别结果而加以输出。 25

在此，使用于语音识别时的上述经杂音适应处理的说话人适应模型 $Mreg$ 如上述般被实行说话人个人特征的适应，并且也被实行对识别说话时的背景杂音的适应。

因此，即使语音识别时的背景杂音环境与适应说话时的背景杂音环

境不同，在语音识别时也可得到很高的识别性能。

接着，参照图 6 的流程图说明本语音识别装置的操作。也就是图 6 揭示说话人适应处理时的操作。

在图 6 中，在开始说话人适应处理时，首先在步骤 S100' 将记忆于
5 初期语音模型记忆部 1 的初期语音模型 M_c 复制写入说话人适应模型记忆部 12，其后由杂音适应部 13 对该初期语音模型 M_c 实施杂音适应处理，以生成杂音适应模型 M_c' 。

也就是，说话人适应处理时的非说话期间所收音的背景杂音的特征
10 矢量系列 $N(n)$ 自声音分析部 16 被供给至杂音适应部，而杂音适应部 13 则借由其特征矢量系列 $N(n)$ 对初期语音模型 M_c 实施杂音适应处理，借以生成杂音适应模型 M_c' ，并送出至说话人适应参数计算部 14。

接着在步骤 S102' 在说话人开始说话时，切换开关 18 切换至说话
人适应参数计算部 14 一侧，而在其说话期间内，叠加有背景杂音的语音
15 (背景杂音叠加语音)的特征矢量系列 $V(n)$ 自语音分析部 16 被供给至说话人适应参数计算部 14。

接着，说话人适应参数计算部 14 借由这些特征矢量系列 $V(n)$ 与杂音适应模型 M_c' 产生说话人适应参数 P 。

也就是，适应已述的 MLLR 或 MAR 算法而求取说话人适应参数 P 时，
将变换行列 W_m' 与偏移矢量 b_m' 作为说话人适应参数 P 加以生成。

20 接着在步骤 S104'，由语音模型更新部 15 使用记忆于说话人适应模型记忆部 12 的初期语音模型 M_c 与说话人适应参数 P 实行模型更新计算，从而求出说话人适应模型 M_c'' 。

接着在步骤 S106' 由语音模型更新部 15 将说话人适应模型 M_c'' 取代记忆于说话人适应模型记忆部 12 的初期语音模型 M_c 作更新记忆后，
25 结束说话人适应处理。

其后，在此说话人适应处理之后，识别处理部 19 在实行语音识别处理之际将更新记忆于说话人适应模型记忆部 12 的说话人适应模型 M_c'' 当作初期语音模型 M_c 利用，而借由将该经更新的初期语音模型 M_c (换句话说就是说话人适应模型 M_c'') 由杂音适应部 13 作杂音适应处理，生成实

5 施过杂音适应处理的说话人适应模型 Mreg 而供给至语音识别部 19,接着,语音识别部 19 对照由前述说话人适应模型 Mreg 所构成的系列与由声音分析部 16 输出的说话人语音的特征矢量系列 $V(n)$ 。接着,将能得到最高优度的说话人适应模型 Mreg 所构成的说话人适应系列当作识别结果作输出。

这样,依本实施方式的语音识别装置,因在说话人适应处理前已实行杂音适应处理,所以在说话人适应处理时所求得的说话人适应参数可减低说话人自适应时的背景杂音的不良影响。

10 还有,因为使用该经减低背景杂音的不良影响的说话人适应参数产生说话人适应模型 Mc'' ,所以可达到说话人适应处理的本来目的,也就是说话人适应处理效果较高的说话人适应模型 Mc'' 。

还有,在语音识别时,将经更新记忆的说话人适应模型 Mc'' 以识别说话时的背景杂音加以杂音适应处理而使用。

15 因此,可使用适应于说话人的个人特征与说话时的背景杂音的两者的模型实行识别,结果可得到很高的识别性能。

(第四实施方式)

接着参照图 7 及图 8 说明本发明的第四实施方式。其中,图 7 揭示本实施方式的语音识别装置的构成的示意图,其与图 5 中相同或相当的部份标注以相同符号。本实施方式为在语音识别处理中实施说话人适应处理。在此,图 7 中所示的讯号的通过路径全部以带箭头的实线表示。

25 在图 7 中,显示了本语音识别装置与第三实施方式的语音识别装置的差异,在第三实施方式的语音识别装置中,其在说话人适应处理后实行语音识别,而本实施方式的语音识别装置在语音识别中同时实行说话人适应处理。

还有,由杂音适应部 13 所输出的杂音适应模型 Mc' 除被送至实行说话人适应处理的说话人适应参数计算部 14 外,说话人适应模型记忆部 12 的内容被说话人适应模型 Mc'' 所更新,而杂音适应模型 Mc' 被作为图 15 所示的经杂音适应处理的说话人适应模型 Mreg 而送至识别处理部 19。

因此，图 7 所示的杂音适应模型 M_c' 自杂音适应部 13 被输出至说话人适应参数计算部 14 以及识别处理部 19，其中送到说话人适应参数计算部 14，作为说话人适应处理用的杂音适应模型 M_c' ，而送到识别处理部 19 作为语音识别处理用的经杂音适应处理的说话人适应模型 M_{reg} 来
5 输出。

识别处理部 19 将已述的说话人适应模型 M_c'' 当作初期语音模型 M_c ，而由杂音适应部 13 对照由杂音适应模型 M_c' （也就是经杂音适应处理的说话人适应模型 M_{reg} ）所构成的系列与应该识别的语音的特征矢量系列 $V(n)$ ，从而将能得到最大优度的说话人适应模型 M_{reg} 所构成的系列当作
10 识别结果来输出。接着，由前述优度产生揭示识别结果与语音的相似性的记录资料 SCR 而与识别结果一同输出。

也就是，实行上述对照的结果，在得到较高优度的情况下，输出表示语音识别结果的可靠度很高的记录资料 SCR 与上述的识别结果，而在无法得到高优度的情况下，输出表示语音识别结果的可靠度较低
15 的记录资料 SCR 与上述的识别结果，而供给至说话人适应参数计算部 14。

之后，当说话人适应参数计算部 14 被供给以表示语音识别结果的可靠度较高的记录资料 SCR 与上述识别结果时，判断为正确地识别出语音，从而产生以该语音识别为对象的语音的特征矢量系列 $V(n)$ ，与杂音适应部 13 所输出的杂音适应模型 M_c' ，与说话人适应处理用的说话人适应参
20 数 P 。

接着，语音模型更新部 15 使用前述说话人适应参数 P 与记忆于说话人适应模型记忆部 12 的初期语音模型 M_c 产生说话人适应模型 M_c'' ，而借由将该说话人适应模型 M_c'' 供给至说话人适应模型记忆部 12，取代语音模型 M_c 而作更新记忆。

因此，本语音识别装置者在处理语音识别处理次数越多时，其记忆于说话人适应模型记忆部 12 的初期语音模型 M_c 的对说话人的个人特征的适应程度逐渐被加以提高。
25

接着参照图 8 所示的流程图说明本语音识别装置的操作。

在图 8 中在开始语音识别处理后，首先在步骤 S200 其杂音适应部

13 对记忆于说话人适应模型记忆部 12 的初期语音模型 M_c 实施杂音适应处理，借以生成杂音适应模型 M_c' 。

也就是，说话人尚未开始说话的前的非说话期间所收音的背景杂音的特征矢量系列 $N(n)$ 自声音分析部 16 被供给至杂音适应部 13，而杂音适应部 13 借由其背景杂音的特征矢量系列 $N(n)$ 将初期语音模型 M_c 作杂音适应处理而产生杂音适应模型 M_c' 。

接着在步骤 S202 在说话人开始说话时，前述切换开关 18 切换至识别处理部 19 一侧，而其说话期间所说话的语音的特征矢量系列 $V(n)$ 自声音分析部 16 被供给至识别处理部 19。

10 接着，识别处理部 19 使用由杂音适应部 13 所产生的杂音适应模型 M_c' 产生识别候补单词模型及识别候补文句模型。

接着，识别处理部 19 在接下来的步骤 S204 中借由对照识别候补单词模型及识别候补文句模型和特征矢量系列 $V(n)$ 而实行语音识别，并输出识别结果与记录资料 SCR。

15 接着，在步骤 S206 前述说话人适应参数计算部 14 判断记录资料 SCR 是否为较高的记录，若不是较高的记录（“No”的情况下），则判断识别结果的可靠度为很低而转移至后述的步骤 S214，而在较高的记录时（“Yes”的情况下）则转移至步骤 S208。

20 在步骤 S208，前述说话人适应参数计算部 14 借由当前识别对象的语音的特征矢量系列 $V(n)$ ，与杂音适应模型 M_c' 及识别结果产生出说话人适应处理用的说话人适应参数 P 。

接着，在步骤 S210 中前述语音模型更新部 15 使用记忆于说话人适应模型记忆部 12 的初期语音模型 M_c 与说话人适应参数 P 实行说话人适应处理，从而求出说话人适应模型 M_c'' 。

25 接着，在步骤 S212 前述语音模型更新部 15 将所生成的说话人适应模型 M_c'' 供给至说话人适应模型记忆部 12，而置换语音模型 M_c ，并且在更新记忆后结束其处理。

这样，依照本实施方式的语音识别装置因为同时进行语音识别与说话人适应处理，所以可产生对说话人的个人特征的适应程度较高的说话

人适应模型 Mc”，从而可更新记忆于说话人适应模型记忆部 12。

因此，通过说出大量的不同的单词及文句而由识别处理部 19 累积性的对这些语音作语音识别，记忆于说话人适应模型记忆部 12 的初期语音模型 Mc 被更新成对说话人个人特征的适应程度较高的说话人适应模型 5 Mc”，借此可达到提高语音识别性能的目的。

还有，在得到高记录的情况下，由于产生说话人适应模型 Mc”而更新初期语音模型 Mc，因此可对应说话环境的状态等实行适当的说话人适应处理，可防止使语音识别性能低下的不适当的说话人适应处理，甚至可实现提高语音识别性能的效果。

10 同时，在同时进行语音识别与说话人适应处理的本实施方式的语音识别装置上，与已述的第三实施方式相同，在以说话人适应处理部实行说话人适应处理前因为以杂音适应部 13 实行杂音适应处理，所以在说话人适应处理时所求取的说话人适应参数 P 具有减低说话人适应处理时的背景杂音的不良影响的优秀效果。

15 如以上所说明，依照本发明的语音识别装置以及语音识别方法，其借由对初期语音模型实施杂音适应处理而产生杂音适应模型，而借由对此杂音适应模型实施说话人适应计算而求取说话人适应参数，而借由对杂音适应处理前的初期语音模型以前述说话人适应参数实施说话人适应处理而产生说话人适应模型，因此可减低说话人适应处理时的背景杂音 20 的不良影响，从而可达到说话人适应处理的本来目的，生成对说话人的个人特征的适应效果较高的说话人适应模型。

还有，在语音识别时因对上述经说话人适应处理的说话人适应模型实施杂音适应处理，而产生实施过杂音适应处理的说话人适应模型，并使用该杂音适应了了的说话人适应模型实行语音识别之处理，由此可使用 25 适应于识别说话时的背景杂音与说话人的个人特征的杂音说话人适应模型而实行语音识别，因此可在各种说话杂音环境下得到很高的识别性能。

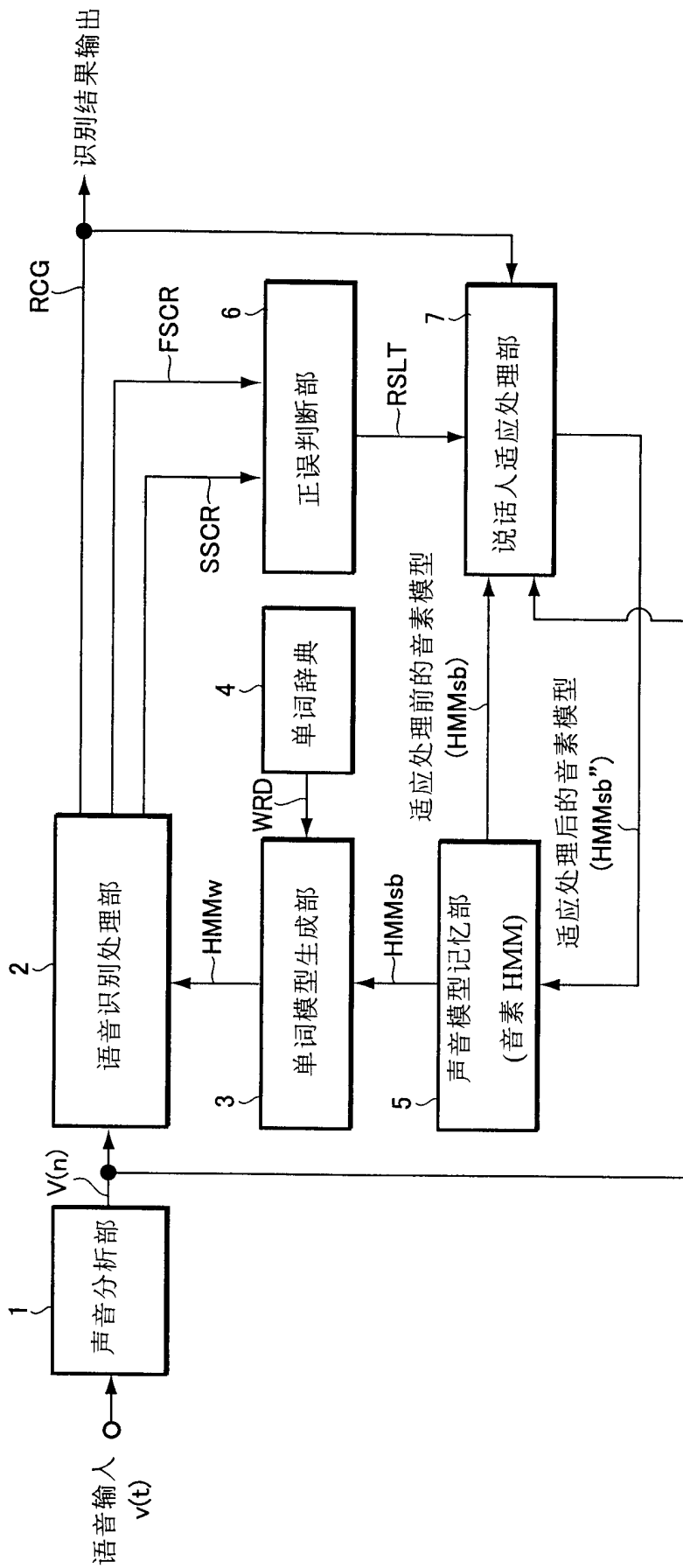


图 1

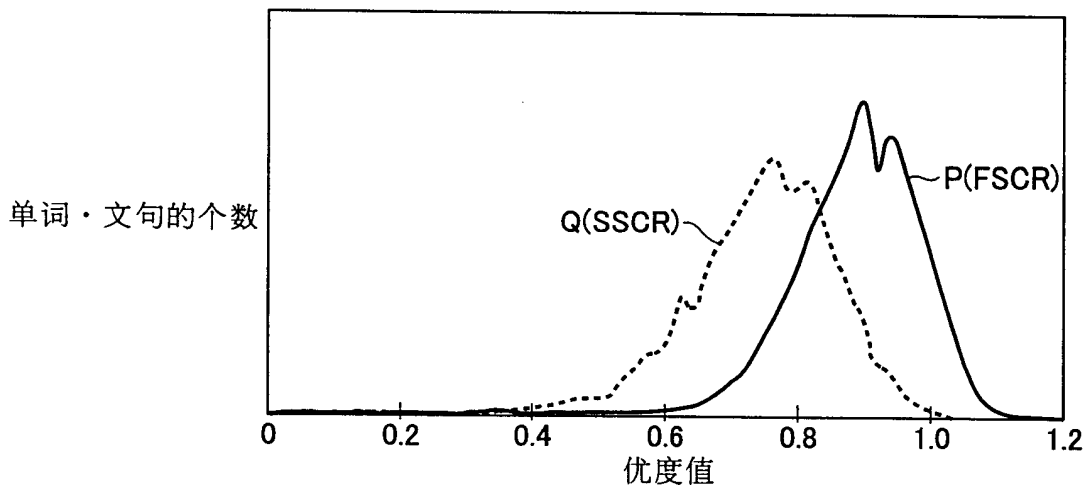


图 2(a)

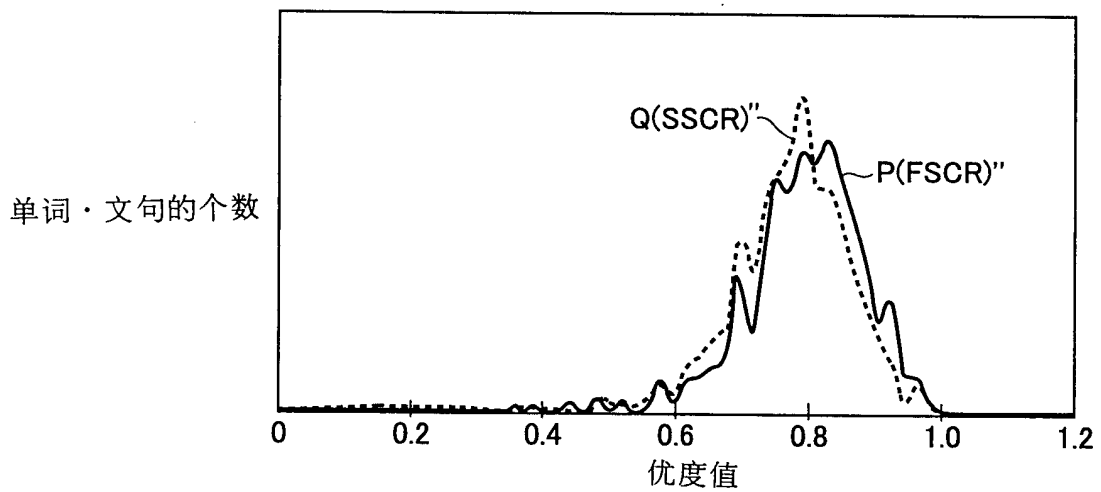


图 2(b)

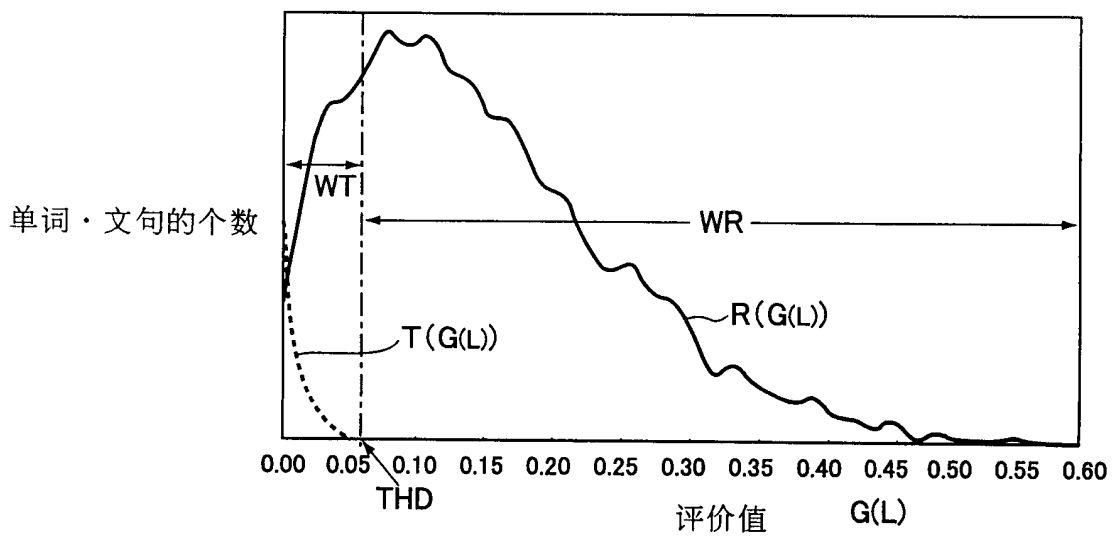


图 2(c)

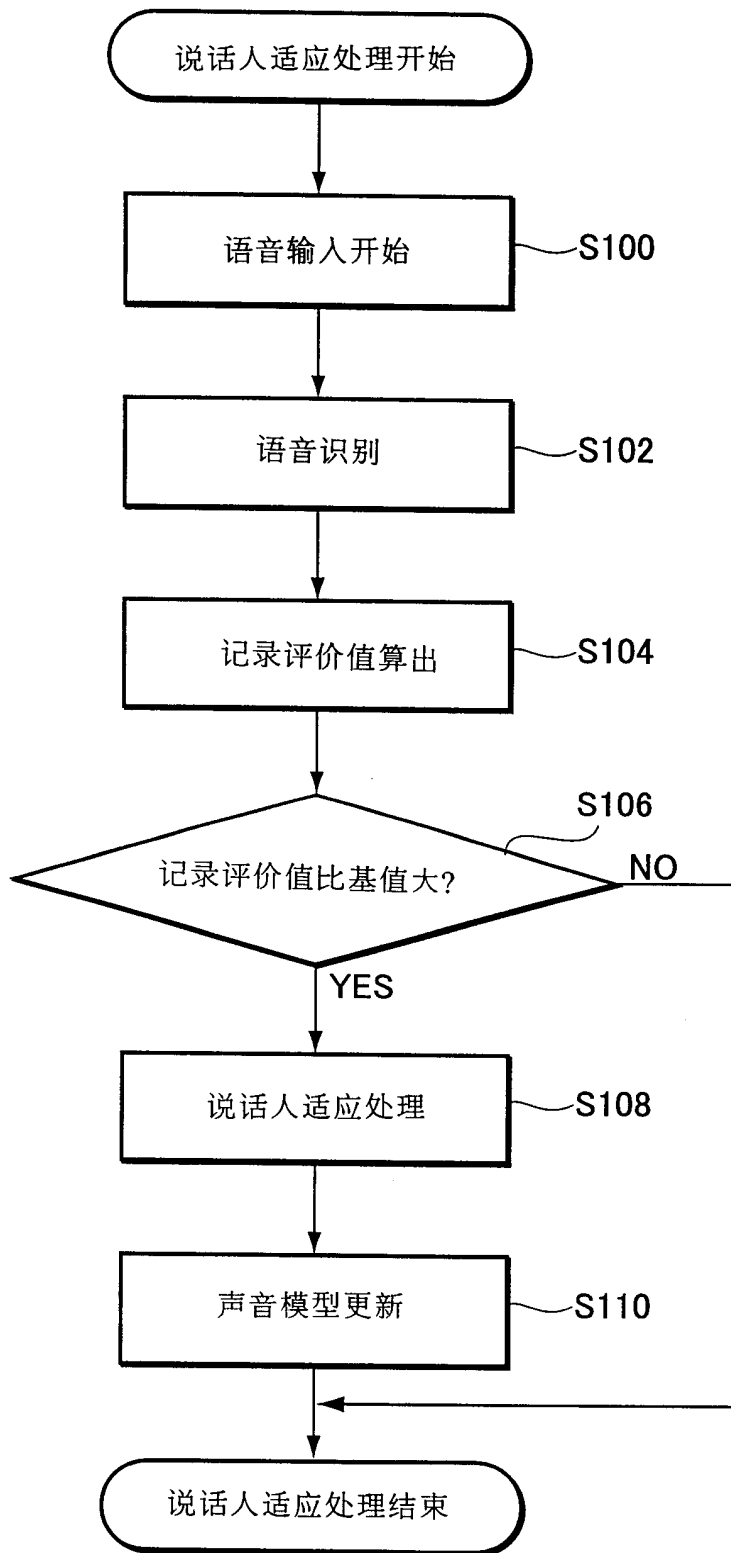


图 3

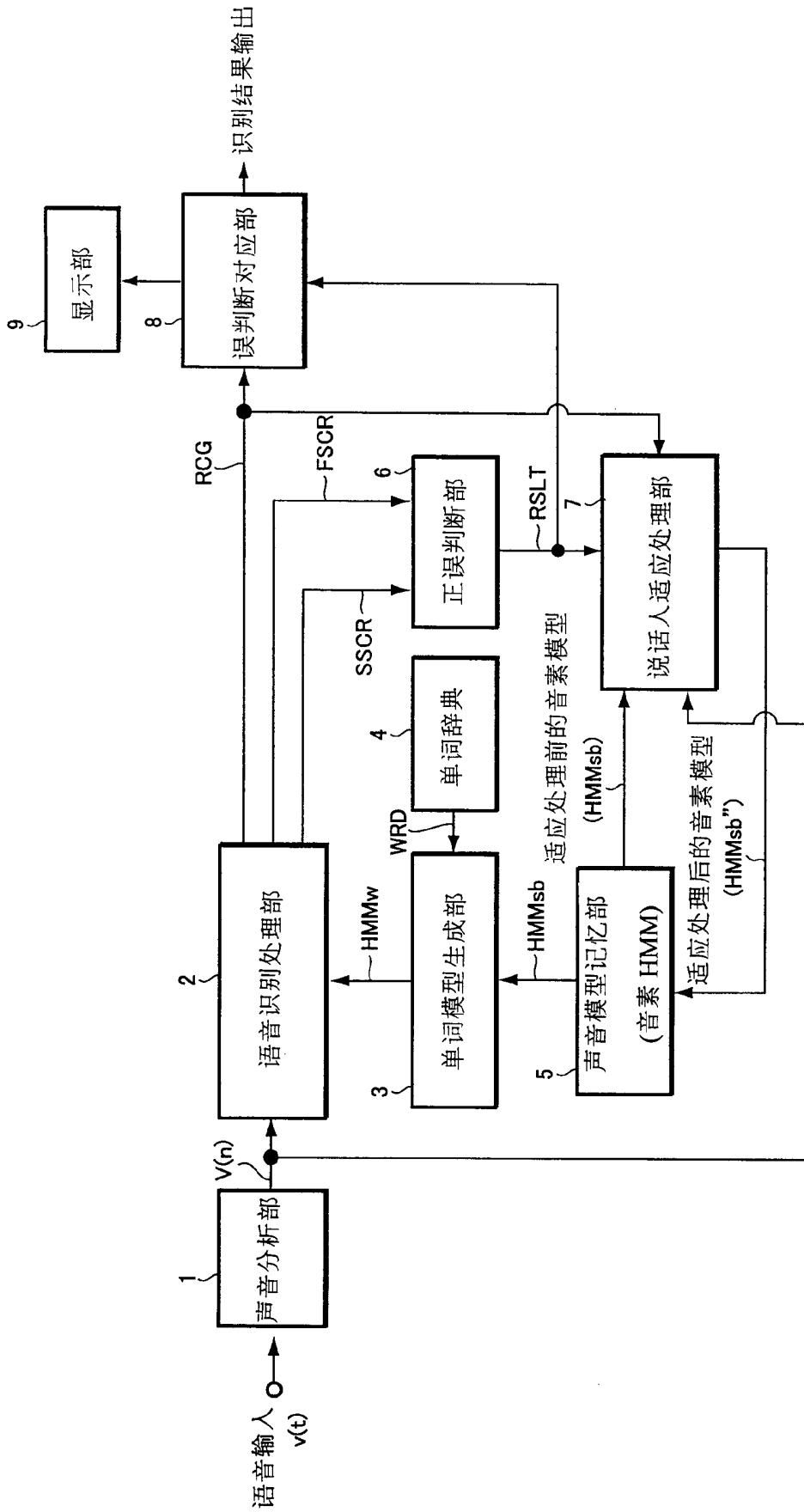


图 4

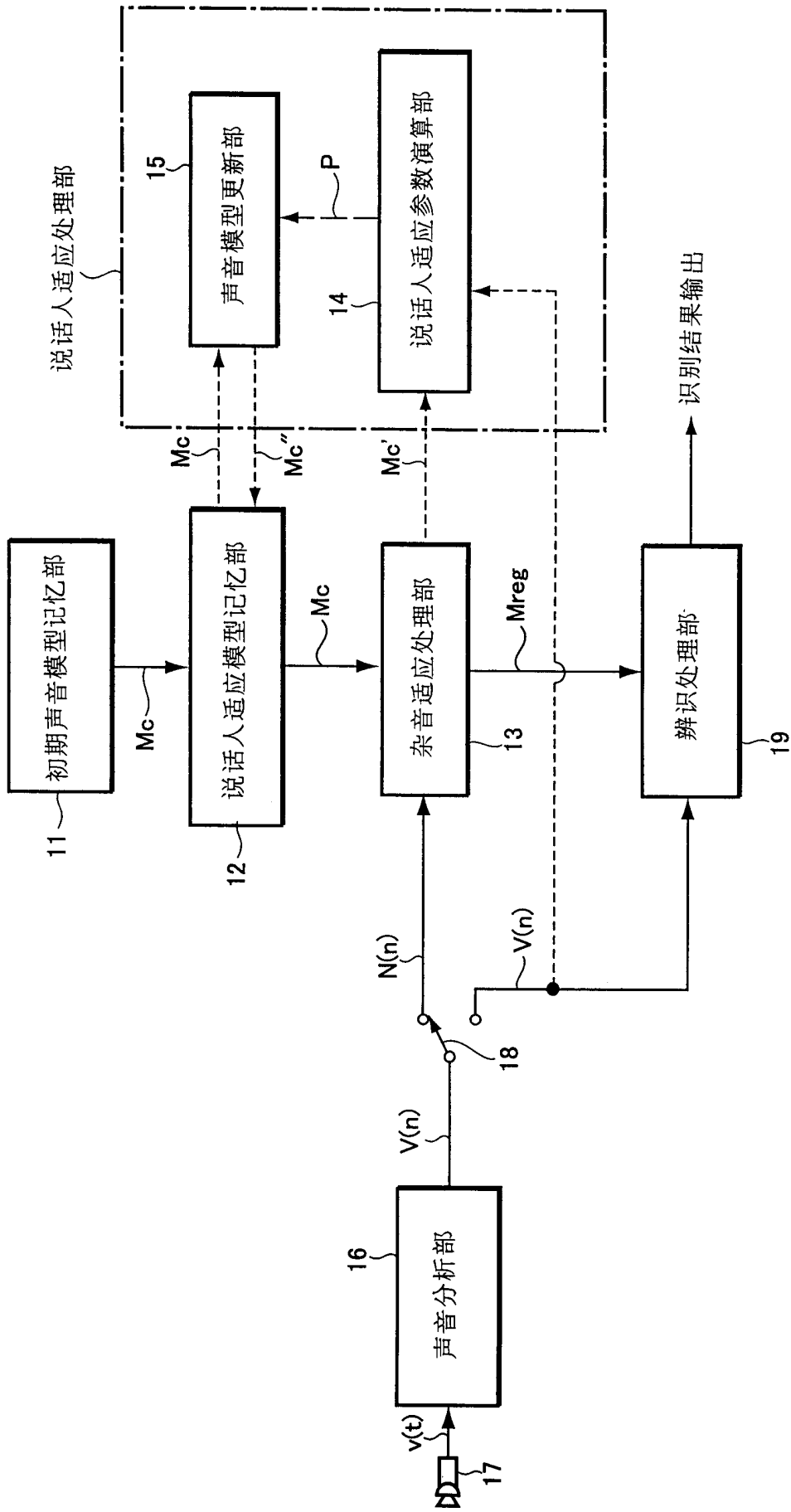


图 5

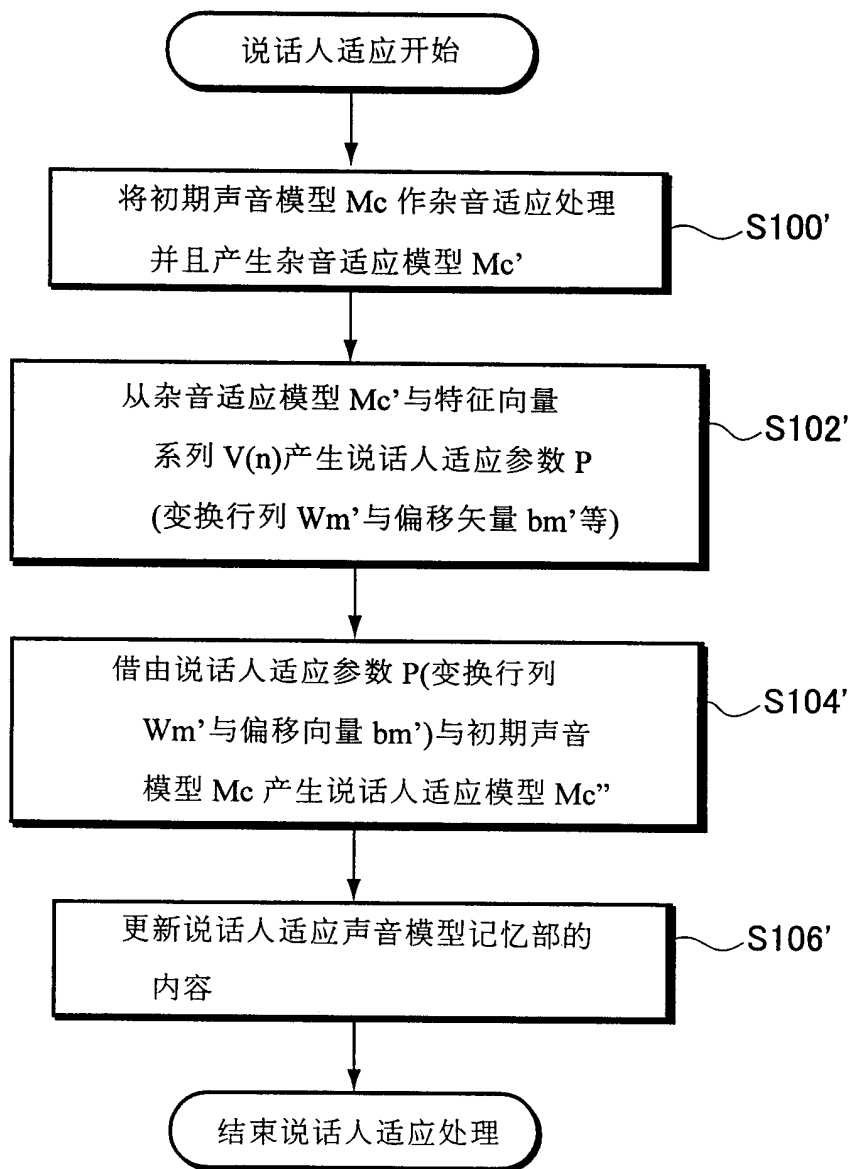


图 6

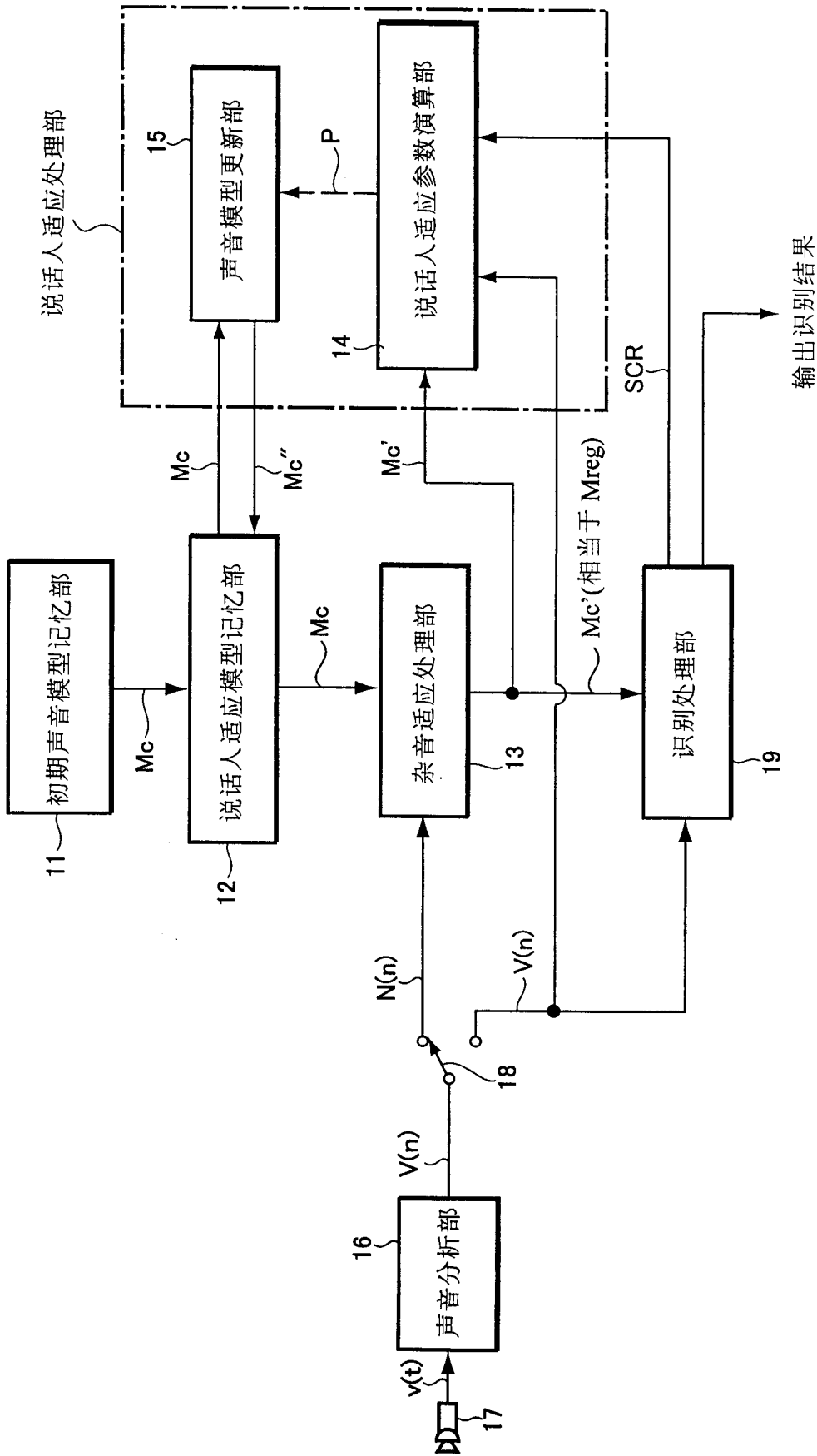


图 7

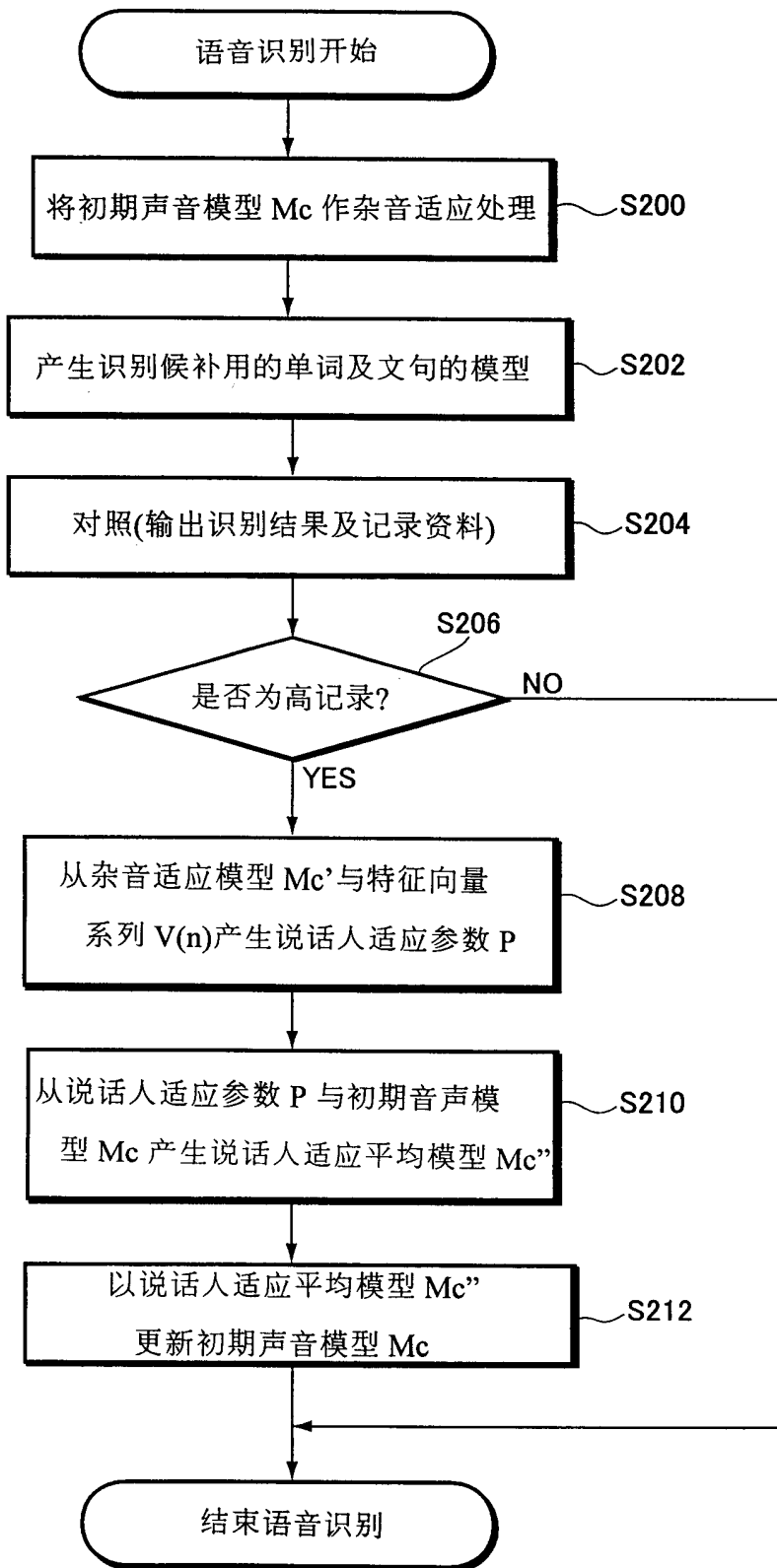


图 8