[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 15/10 (2006.01)
G10L 11/00 (2006.01)
G10L 15/00 (2006.01)
G10L 11/06 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200780000900.4

[43] 公开日 2009 年 1 月 14 日

[11] 公开号 CN 101346758A

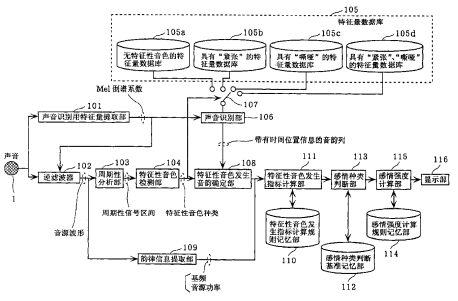
[22] 申请日 2007.5.21
[21] 申请号 200780000900.4
[30] 优先权
[32] 2006.6.23 [33] JP [31] 173937/2006
[86] 国际申请 PCT/JP2007/060329 2007.5.21
[87] 国际公布 WO2007/148493 日 2007.12.27
[85] 进入国家阶段日期 2008.2.29
[71] 申请人 松下电器产业株式会社
地址 日本大阪府
[72] 发明人 加藤弓子 釜井孝浩 中藤良久 广濑良文
[74] 专利代理机构 永新专利商标代理有限公司
代理人 胡建新 杨 谦
权利要求书 5 页 说明书 44 页 附图 20 页

[54] 发明名称

感情识别装置

[57] 摘要

提供一种感情识别装置，与音韵信息的个人差别、地方差别、依据语言的差别无关，可以准确且稳定地进行依据声音的感情识别，所述感情识别装置依据输入声音来识别该输入声音的讲话者的感情，包括：特征性音色检测单元，从所述输入声音中检测与特定的感情有关的特征性音色；声音识别单元(106)，识别输入声音中包含的音韵的种类；特征性音色检测单元(104)，从所述输入声音中检测与特定的感情有关的特征性音色；特征性音色发生指标计算单元(111)，根据所述声音识别单元(106)所识别的音韵的种类，按每个音韵计算特征性音色发生指标，该特征性音色发生指标示出所述特征性音色的发生容易度；以及感情判断单元(113)，根据所述特征性音色发生指标计算单元(111)所计算的所述特征性音色发生指标来判断在发生了所述特征性音色的音韵中的所述输入声音的讲话者的感情。



1、一种感情识别装置，依据输入声音来识别该输入声音的讲话者的感情，其特征在于，包括：

特征性音色检测单元，从所述输入声音中检测与特定的感情有关的特征性音色；

声音识别单元，根据所述特征性音色检测单元检测出的特征性音色来识别所述输入声音中包含的音韵的种类；

特征性音色发生指标计算单元，根据所述声音识别单元所识别的音韵的种类，按每个音韵计算特征性音色发生指标，该特征性音色发生指标示出所述特征性音色的发生容易度；以及

感情判断单元，根据规则，依据所述特征性音色发生指标计算单元所计算的所述特征性音色发生指标来判断在发生了所述特征性音色的音韵中的所述输入声音的讲话者的感情，所述规则就是所述特征性音色发生指标越小所述感情越强。

2、如权利要求 1 所述的感情识别装置，其特征在于，还包括：

感情强度判别单元，根据计算规则，来判别发生了所述特征性音色的音韵中的感情强度，所述计算规则是指所述特征性音色发生指标越小所述感情强度越强。

3、如权利要求 2 所述的感情识别装置，其特征在于，

所述感情强度判别单元，对所述特征性音色发生指标计算单元所计算的每个音韵的特征性音色发生指标、和所述特征性音色检测单元

检测出的发生了特征性音色的时间上的声音位置进行比较，根据计算规则，来判别发生了所述特征性音色的音韵中的感情强度，所述计算规则是指特征性音色发生指标越小所述感情强度越强。

4、如权利要求 1 所述的感情识别装置，其特征在于，

所述特征性音色检测单元，将在母音部位的音源存在微扰的声音的音色作为特征性音色检测。

5、如权利要求 1 所述的感情识别装置，其特征在于，还至少包括：

特征量数据库，按每个音韵的种类记忆包含所述特征性音色的声音的特征量，

所述声音识别单元，根据所述特征量数据库，来识别所述输入声音中包含的音韵的种类。

6、如权利要求 5 所述的感情识别装置，其特征在于，

所述特征量数据库，包括：

第一数据库，按每个所述音韵的种类记忆包含至少一个以上的所述特征性音色的声音的特征量；以及

第二数据库，按每个所述音韵的种类记忆不包含所述特征性音色的声音的特征量，

所述感情识别装置，还包括选择单元，从所述第一数据库以及所述第二数据库中选择，与所述特征性音色检测单元中的检测结果相对应的数据库，

所述声音识别单元，根据所述选择单元所选择的数据库，来识别所述输入声音中包含的音韵的种类。

7、如权利要求1所述的感情识别装置，其特征在于，还包括：

音响特征量数据库，按每个音韵的种类记忆音响特征量；以及

语言特征量数据库，包含表示单词辞典的语言特征量，该单词辞典至少具有读法或发音记号，

所述声音识别单元，对于检测出所述特征性音色的单词，将所述音响特征量数据库中包含的音响特征量的权重变小、将所述语言特征量数据库中包含的语言特征量的权重变大，并根据所述音响特征量数据库和所述语言特征量数据库，从而识别所述输入声音中包含的音韵的种类。

8、一种感情识别装置，依据输入声音来识别该输入声音的讲话者的感情，其特征在于，包括：

特征性音色检测单元，从所述输入声音中检测与特定的感情有关的特征性音色；

音韵输入单元，将输入声音中包含的音韵的种类输入；

特征性音色发生指标计算单元，至少将从所述音韵输入单元输入的音韵的种类作为参数利用，从而按每个音韵计算特征性音色发生指标，该特征性音色发生指标示出所述特征性音色的发生容易度；以及

感情判断单元，根据对应规则，依据所述特征性音色发生指标计算单元所计算的所述特征性音色发生指标来判断在发生了所述特征性音色的音韵中的所述输入声音的讲话者的感情，所述对应规则就是所述特征性音色发生指标越小所述感情越强。

9、一种感情识别装置，依据输入声音来识别该输入声音的讲话者

的感情，其特征在于，包括：

声音识别单元，识别输入声音中包含的音韵的种类；

特征性音色检测单元，从所述输入声音中提取在母音部位的音源存在振幅微扰或频率微扰的紧张声音部位；以及

感情判断单元，按每个所述声音识别单元所识别的音韵，在所述特征性音色检测单元检测出的声音部位是包括至少一个下列音的音韵的情况下，判断为所述输入声音的讲话者的感情是强烈的愤怒，所述音是：由嘴唇构音的无声爆破音；由牙齿构音的无声破擦音；由嘴唇和牙齿构音的无声摩擦音。

10、一种感情识别方法，依据输入声音来识别该输入声音的讲话者的感情，其特征在于，包括：

特征性音色检测步骤，从所述输入声音中检测与特定的感情有关的特征性音色；

声音识别步骤，根据所述特征性音色检测步骤检测出的特征性音色来识别所述输入声音中包含的音韵的种类；

特征性音色发生指标计算步骤，根据所述声音识别步骤所识别的音韵的种类，按每个音韵计算特征性音色发生指标，该特征性音色发生指标示出所述特征性音色的发生容易度；以及

感情判断步骤，根据规则，依据所述特征性音色发生指标计算步骤所计算的所述特征性音色发生指标来判断在发生了所述特征性音色的音韵中的所述输入声音的讲话者的感情，所述规则就是所述特征性音色发生指标越小所述感情越强。

11、如权利要求 10 所述的感情识别方法，其特征在于，还包括：

感情强度判别步骤，根据计算规则，来判别发生了所述特征性音色的音韵中的感情强度，所述计算规则是指所述特征性音色发生指标越小所述感情强度越强。

12、一种程序，依据输入声音来识别该输入声音的讲话者的感情，其特征在于，使计算机执行以下步骤：

特征性音色检测步骤，从所述输入声音中检测与特定的感情有关的特征性音色；

声音识别步骤，根据所述特征性音色检测步骤检测出的特征性音色来识别所述输入声音中包含的音韵的种类；

特征性音色发生指标计算步骤，根据所述声音识别步骤所识别的音韵的种类，按每个音韵计算特征性音色发生指标，该特征性音色发生指标示出所述特征性音色的发生容易度；以及

感情判断步骤，根据规则，依据所述特征性音色发生指标计算步骤所计算的所述特征性音色发生指标来判断在发生了所述特征性音色的音韵中的所述输入声音的讲话者的感情，所述规则就是所述特征性音色发生指标越小所述感情越强。

13、如权利要求 12 所述的程序，其特征在于，进一步，使计算机执行：

感情强度判别步骤，根据计算规则，来判别发生了所述特征性音色的音韵中的感情强度，所述计算规则是指所述特征性音色发生指标越小所述感情强度越强。

感情识别装置

技术领域

本发明涉及依据声音来识别讲话者的感情的感情识别装置。进一步，确定而言，尤其涉及一种依据声音的感情识别装置，依据因讲话者的感情、表情、态度或讲话风格而经常变化的发声器官的紧张或松弛，来识别在发声的声音中是否发生了特征性音色，从而识别讲话者的感情。

背景技术

在自动电话对应、电子秘书、对话机器人等具有依据声音对话的接口的对话系统中，为了对话系统依据用户的请求来适当地进行对应，重要的条件是依据用户发声的声音来理解用户的感情。例如，在所述的自动电话对应或对话机器人与用户进行依据声音的对话时，在对话系统的声音识别中不一定可以准确地识别声音。在对话系统产生识别错误的情况下，对话系统向用户再次请求输入声音。在这些状况下，用户多少也觉得愤怒、急躁。若多次产生识别错误，则用户越发觉得愤怒、急躁。愤怒或急躁，使用户的说法或声质变化，而用户的声音成为与平时的声音不同的模式。因此，在将平时的声音作为识别用模型保持的对话系统会更容易产生识别错误，因此向用户多次请求相同

回答等进行对用户更不愉快的请求。在对话系统处于如上所述的恶循环的情况下，其不能用以对话接口。

为了停止这些恶循环、使机器和用户的声音对话正常化，需要依据用户发声的声音来识别感情。即，若可以理解用户的愤怒或急躁，则在识别错误时，对话系统可以以更礼貌的口气来再问或道歉。据此，对话系统可以使用户的感情接近平常状态、而引起平常的讲话，从而恢复识别率。甚至，可以顺利地进行依据对话系统的机器操作。

在以往的技术中，作为依据声音来识别感情的方法提出了下列方式：从讲话者发声的声音中提取声音的高度(基频)、大小(功率)、讲话速度等韵律特征，对输入声音整体进行根据像“声音高”、“声音大”那样的判断的感情识别(例如，参照专利文献 1、专利文献 2)。而且，提出了下列方式：对输入声音整体进行像“高频域的能量大”那样的判断(例如，参照专利文献 1)。还提出了下列方式：依据声音的功率和基频的序列来求出像它们的平均、最大值、最少值那样的统计上的代表值，从而识别感情(例如，参照专利文献 3)。进一步，还提出了下列方式：利用像句子、单词的语调或声调(accent)那样的韵律的时间模式来识别感情(例如，参照专利文献 4、专利文献 5)。

图 20 是所述专利文献 1 所记载的、以往的依据声音的感情识别装置的示意图。

麦克风 1 将输入声音转换为电信号。声音码识别单元 2，对麦克风 1 所输入的声音进行声音识别，将识别结果输出到感情信息提取单元 3 以及输出控制单元 4。

另外，感情信息提取单元 3 的话速检测部 31、基频检测部 32 以及音量检测部 33，分别从麦克风 1 所输入的声音中提取话速、基频以及音量。

声音级判断基准存储部 34 记忆有用于决定声音级的基准，该声音级是，将所输入的声音的话速、基频以及音量分别与标准的话速、基频以及音量比较来决定的。标准声音特征量存储部 35 记忆有，成为判断声音级时的基准的、标准的发声速度、基频以及音量。声音级分析部 36，根据所输入的声音的特性量和标准的声音特征量的比率来决定声音级，即，决定话速级、基频级以及音量级。

进一步，感性级分析用知识库存储部 37 记忆有规则，该规则用于依据声音级分析部 36 所决定的各种声音级来判断感性级。感性级分析部 38，依据来自声音级分析部 36 的输出和来自声音码识别单元 2 的输出，并根据感性级分析用知识库存储部 37 记忆有的规则，从而判断感性级，即，判断感性的种类和级。

输出控制单元 4，根据感性级分析部 38 所输出的感性级来控制输出装置 5，从而生成与所输入的声音的感性级相对应的输出。在此，用于决定声音级的信息是，以表示在每一秒中讲了几个音拍的话速、平均基频、讲话、句子或短句那样的单位来求出的韵律信息。

然而，韵律信息的特点是，韵律信息也用于表达语言信息，而且该语言信息的表达方法，按每个语言种类不同。例如，在日语中，存在像“はし(桥)”和“はし(筷子)”那样的多个同音异义词，根据由基频的高低作出的声调词汇的意思不同。而且，在汉语中，已知按照叫

做四声声调的基频的变动，即使同音也会表示完全不同的意思(文字)。在英语中，不依据基频而依据叫做重读(stress)的声音的强度来表示声调，并且，重读的位置成为在区别单词及句子的意思、或词类时的线索。为了进行依据韵律的感情识别，需要考虑这些依据语言的韵律模式的差别，按每个语言分离作为感情表现的韵律的变化和作为语言信息韵律的变化，来生成感情识别用数据。而且，即使在同一语言内，也在利用韵律的感情识别中存在说得快的人、声音高(低)的人等个人差别，因此，例如，平时以大声讲话的、说得快且声音高的人，经常被识别为生气。因此，还需要下列方法：记忆每个人的标准数据，通过按每个人与标准数据比较来进行符合每个人的感情识别，从而防止依据个人差别的识别错误(例如，专利文献 2、专利文献 5)。

专利文献 1：日本国特开平 9-22296 号公报(第 6-9 页，表 1-5，第 2 图)

专利文献 2：日本国特开 2001-83984 号公报(第 4-5 页，第 4 图)

专利文献 3：日本国特开 2003-99084 号公报

专利文献 4：日本国特开 2005-39501 号公报(第 12 页)

专利文献 5：日本国特开 2005-283647 号公报

如上所述，在依据韵律的感情识别中，由于按每个语言分离韵律信息中用于表示语言信息的变动和作为感情表现的变动，因此需要大量的声音数据、分析处理以及统计处理。再者，即使在同一语言内也地方差别或依据年龄等的个人差别大，并且，即使同一讲话者也依据身体状态等的变动大。因此，在不保存每个用户的标准数据的情况下，

依据韵律的感情表现的地方差别或个人差别大，因此对非特定多数的声音难以不断地生成稳定的结果。

再者，对于保存每个人的标准数据的方式，在估计非特定多数的人使用的呼叫中心或车站等公共场所的向导系统等不能采用。这是因为，不能保存每个讲话者的标准数据。

而且，对于韵律数据，需要以讲话、句子、短句等声音的一定长度，对每一秒的音拍数、平均及动态范围等统计性代表值、或时间模式等进行分析。据此，存在的课题是，在短时间内声音的特征变化的情况下，难以进行跟随分析，不能进行高精度的依据声音的感情识别。

发明内容

为了解决所述以往的课题，本发明的目的在于，提供依据声音的感情识别装置，以音韵单位那样的较短的单位可以检测感情，并且，利用个人差别、语言差别以及地方差别比较少的特征性音色、和与讲话者的感情的关系，从而进行高精度的感情识别。

本发明涉及的感情识别装置，依据输入声音来识别该输入声音的讲话者的感情，其特征在于，包括：特征性音色检测单元，从所述输入声音中检测与特定的感情有关的特征性音色；声音识别单元，根据所述特征性音色检测单元检测出的特征性音色来识别所述输入声音中包含的音韵的种类；特征性音色发生指标计算单元，根据所述声音识别单元所识别的音韵的种类，按每个音韵计算特征性音色发生指标，该特征性音色发生指标示出所述特征性音色的发生容易度；以及感情

判断单元，根据规则，依据所述特征性音色发生指标计算单元所计算的所述特征性音色发生指标来判断在发生了所述特征性音色的音韵中的所述输入声音的讲话者的感情，所述规则就是所述特征性音色发生指标越小所述感情越强。

声音的物理特征的发生机制是依据下列发声器官的生理原因来求出的：例如爆破音，用嘴唇、舌、腭临时堵塞声道后一下子开放，由于该动作嘴唇或舌容易紧张。据此，可以检测出在依据讲话者的感情或讲话态度来发声器官紧张或松驰而在声音的每一个细节中观察到的、像假声及紧张的声音或气息性声音那样的特征性音色。根据该特征性音色的检测结果，在不受语言种类的差别、依据讲话者的特性的个人差别以及地方差别的影响的情况下，可以以音韵为单位来识别讲话者的感情。

优选的是，所述感情识别装置还包括，感情强度判别单元，根据计算规则，来判别发生了所述特征性音色的音韵中的感情强度，所述计算规则是指所述特征性音色发生指标越小所述感情强度越强。

而且，所述感情强度判别单元，对所述特征性音色发生指标计算单元所计算的每个音韵的特征性音色发生指标、和所述特征性音色检测单元检测出的发生了特征性音色的时间上的声音位置进行比较，根据计算规则，来判别发生了所述特征性音色的音韵中的感情强度，所述计算规则是指特征性音色发生指标越小所述感情强度越强。

在不易发生特征性音色的音韵发生了特征性音色的情况下，可以认为与该特征性音色相对应的特定的感情强。因此，根据这些规则，

在不受语言差别、个人差别以及地方差别的影响的情况下，可以准确地判别感情的强度。

优选的是，所述感情识别装置还包括：音响特征量数据库，按每个音韵的种类记忆音响特征量；以及语言特征量数据库，包含表示单词辞典的语言特征量，该单词辞典至少具有读法或发音记号，所述声音识别单元，对于检测出所述特征性音色的单词，将所述音响特征量数据库中包含的音响特征量的权重变小、将所述语言特征量数据库中包含的语言特征量的权重变大，并根据所述音响特征量数据库和所述语言特征量数据库，从而识别所述输入声音中包含的音韵的种类。

对于发生了特征性音色的单词，将语言特征量的权重变大，从而可以防止因在特征性音色的发生位置不符合音响特征量而引起的识别精度的降低。据此，可以准确地识别感情。

并且，本发明，除了可以作为包括如上所述的特征性单元的感情识别装置来实现以外，也可以作为将感情识别装置中包括的特征性单元作为步骤的感情识别方法来实现，还可以作为使计算机执行感情识别方法中包括的特征性步骤的程序来实现。并且，当然也可以通过CD-ROM(Compact Disc-Read Only Memory)等存储介质或互联网等通信网络来分发这些程序。

根据本发明的依据声音的感情识别装置，可以检测出具有相当于异常值的特性的特征性音色，该异常值示出下列值：在依据讲话者的感情或讲话态度来发声器官紧张或松驰而引起的、脱离了平均的讲话(以平常发声的讲话)的状态的讲话状态中，即在声音的每一个细节中观

察到的、像假声及紧张的声音或气息性声音那样的特定的音响特性中，远离平均的发声的值。通过利用所述特征性音色的检测结果，在不受语言种类、依据讲话者的特性的个人差别以及地方差别的影响的情况下，可以以音韵为单位来识别讲话者的感情，因此可以跟随讲话中的感情变化。

附图说明

图 1A 是针对讲话者 1 按每个音拍内的子音示出，以带有“强烈的愤怒”的感情表现的声音中的“紧张的”音或“刺耳的声音（harsh voice）”发声的音拍的频度的图表。

图 1B 是针对讲话者 2 按每个音拍内的子音示出，以带有“强烈的愤怒”的感情表现的声音中的“紧张的”音或“刺耳的声音(harsh voice）”发声的音拍的频度的图表。

图 1C 是针对讲话者 1 按每个音拍内的子音示出，以带有“中等程度的愤怒”的感情表现的声音中的“紧张的”音或“刺耳的声音(harsh voice）”发声的音拍的频度的图表。

图 1D 是针对讲话者 2 按每个音拍内的子音示出，以带有“中等程度的愤怒”的感情表现的声音中的“紧张的”音或“刺耳的声音(harsh voice）”发声的音拍的频度的图表。

图 2A 是针对讲话者 1 示出，录音的声音中的特征性音色“嘶哑”的声音的、依据音韵种类的发生频度的图表。

图 2B 是针对讲话者 2 示出，录音的声音中的特征性音色“嘶哑”

的声音的、依据音韵种类的发生频度的图表。

图 3A 是录音的声音中观察到的特征性音色的声音的发生位置、和推定的特征性音色的声音的时间位置的比较图。

图 3B 是录音的声音中观察到的特征性音色的声音的发生位置、和推定的特征性音色的声音的时间位置的比较图。

图 4 是本发明的实施例 1 的依据声音的感情识别装置的框图。

图 5 是本发明的实施例 1 的依据声音的感情识别装置的工作流程图。

图 6 是示出本发明的实施例 1 的特征性音色发生指标的计算规则的一个例子的图。

图 7 是示出本发明的实施例 1 的感情种类判断规则的一个例子的图。

图 8 是示出本发明的实施例 1 的感情强度计算规则的一个例子的图。

图 9 是有“紧张”的音拍的发生频度和没有“紧张”的音拍的发生频度和指标的值的关系、以及感情的强度(弱度)和指标的值的关系的模式图。

图 10 是本发明的实施例 1 的变形例的依据声音的感情识别装置的框图。

图 11 是本发明的实施例 1 的变形例的依据声音的感情识别装置的工作流程图。

图 12 是录音的声音中观察到的特征性音色的声音的发生位置、和

该特征性音色的发生容易度的比较图。

图13是示出本发明的实施例1的变形例的感情种类判断规则的一个例子的图。

图14是本发明的实施例2的依据声音的感情识别装置的框图。

图15是本发明的实施例2的依据声音的感情识别装置的工作流程图。

图16A是示出本发明的实施例2的声音识别处理的具体例子的图。

图16B是示出本发明的实施例2的声音识别处理的具体例子的图。

图16C是示出本发明的实施例2的声音识别处理的具体例子的图。

图 17 是本发明的实施例 3 的依据声音的感情识别装置的功能框图。

图18是本发明的实施例3的感情识别装置的工作流程图。

图19是示出本发明的实施例3的音韵输入方法的一个例子的图。

图20是以往的依据声音的感情识别装置的框图。

符号说明

1 麦克风

2 声音码识别单元

3 感性信息提取单元

4 输出控制单元

5 输出装置

31 话速检测部

32 基频检测部

33 音量检测部

34 声音级确定基准存储部

35 标准声音特征量存储部

36 声音级分析部

37 感性级分析用知识库存储部

38 感性级分析部

101 声音识别用特征量提取部

102 逆滤波器

103 周期性分析部

104 特征性音色检测部

105 特征量数据库

106 声音识别部

107 开关

108 特征性音色发生音韵确定部

109 韵律信息提取部

110 特征性音色发生指标计算规则记忆部

111 特征性音色发生指标计算部

112 感情种类判断基准记忆部

113 感情种类判断部

114 感情强度计算规则记忆部

115 感情强度计算部

116 显示部

132 感情种类判断规则记忆部

133 感情种类强度计算部

205 音响特征量数据库

206 语言特征量数据库

207 连续单词声音识别部

208 特征性音色发生音韵确定部

具体实施例

首先，对于成为本发明的基础的、声音中的特征性音色和讲话者的感情的关系，说明在声音中实际上出现的现象。

在带有感情或表情的声音中，已知有各种各样音质的声音混入，因此，对声音的感情或表情赋予了特征，形成了声音的印象（例如，参照：粕谷英樹、楊長盛，“音源から見た声質”，日本音響学会誌 51 卷 11 号（1995），pp869-875；日本国特开 2004-279436 号公报)。在本发明申请之前，对基于相同文本讲话的 50 句进行了无表情的声音与带有表情的声音的调查。

图 1A 是，针对讲话者 1 按每个音拍内的子音示出以带有“强烈的愤怒”的感情表现的声音中的“紧张的”音或“刺耳的声音(harsh voice)”发声的音拍的频度的图表。图 1B 是，针对讲话者 2 按每个音拍内的子音示出以带有“强烈的愤怒”的感情表现的声音中的“紧张的”音或“刺耳的声音(harsh voice)”发声的音拍的频度的图表。图 1C 以及图

1D 是，分别针对与图 1A 及图 1B 相同的讲话者，按每个音拍内的子音示出带有“中等程度的愤怒”的感情表现的声音中的“紧张的”音或“刺耳的声音(harsh voice)”的音拍的频度的图表。

特征性音色的发生频率根据子音的种类而有偏差，对于图 1A 以及图 1B 的图表中所示的各个讲话者，可以发现两名讲话者共通的下列特征：在“t”(由硬腭构音的无声爆破子音)、“k”(由软腭构音的无声爆破子音)、“d”(由硬腭构音的有声爆破子音)、“m”(由嘴唇构音的鼻音)、“n”(由硬腭构音的鼻音)、或无子音的情况下发生频度高，在“p”(由嘴唇构音的无声爆破音)、“ch”(由牙齿构音的无声破擦音)、“ts”(无声破擦音)、“f”(由嘴唇和牙齿构音的无声摩擦音)等发生频度低。即，图 1A 以及图 1B 的图表示出，“愤怒”的感情的声音中出现的“紧张”的发生条件是讲话者间共通的。图 1A 以及图 1B 所示的两名讲话者的“紧张”的发生的偏差倾向，是根据该音拍的子音的种类相同的。而且，即使带着相同程度的“愤怒”的感情而讲话的声音，也根据音韵的种类以“紧张的”音发声的概率不同，若以“紧张的”音发声的概率更低的种类的音韵中检测出以“紧张的”音的发声，则可以推测为“愤怒”的感情的程度大。

而且，对示出作为同一人物的讲话者 1 的特征性音色“紧张”的出现频度的图 1A 和图 1C 进行比较。有像“sh”或“f”那样的种类，在图 1C 所示的中等程度的愤怒的表现中不发生“紧张的”音，但是，在图 1A 所示的强烈的愤怒的表现中发生“紧张的”音。而且，也有像无子音的音拍那样的种类，在图 1C 所示的中等程度的愤怒的表现中

“紧张的”音的发生频度低，但是，在图 1A 所示的强烈的愤怒的表现中“紧张的”音的发生频度增大。如此，可以了解的是，若愤怒的强度变强，在本来不易紧张的音韵中也会发生“紧张的”音。再者，如对讲话者 1 和讲话者 2 已确认，以“紧张的”音发声的每个音韵的偏差是讲话者间共通的。

图 2A 以及图 2B 是，按每个音拍内的子音示出以带有“快活”的感情表现的声音中的“气息性”发声的、即以“嘶哑”或“温和的声音（soft voice)”发声的音拍的频度的图表。图 2A 以及图 2B 是，分别针对讲话者 1 和讲话者 2，按每个音拍内的子音示出以带有“快活”的感情表现的声音中的“气息性”发声的、即以“嘶哑”或“温和的声音（soft voice)”发声的音拍的频度的图表。特征性音色的发生频率根据子音的种类而有偏差，对于图 2A 以及图 2B 的图表中所示的各个讲话者，可以发现两名讲话者共通的下列特征：在“h”(由声门构音的无声摩擦子音)、“k”(由软腭构音的无声爆破子音)的情况下发生频度高，“d”(由硬腭构音的有声爆破子音)、“m”(由嘴唇构音的鼻音)、“g”(由硬腭构音的有声爆破子音)等发生频度低。而且，对于图 2A 以及图 2B 的“b”、“g”、“m”的音韵中的特征性音色的发生频度，讲话者 1 的发生频度是 0，讲话者 2 的发生频度虽然低也存在。一个讲话者的发生频度是 0、另一个讲话者的发生频度虽然低也存在，这些倾向与图 1A～图 1D 中的“f”的音韵的倾向(图 1A 的讲话者 1 的发生频度低、图 1B 的讲话者 2 的发生频度是 0)相同。因此，可以认为的是，图 1A～图 1D 的“f”是本来不易紧张的音韵，也是愤怒的强度变强时发生的音韵，

与此相同，图 2A 以及图 2B 的“b”、“g”、“m”的“嘶哑”音是本来不易嘶哑的音韵，也是“快活”的强度变强时发生的音韵。

如上所述的、依据音韵的发生概率的偏差和讲话者间共通的偏差是，在“紧张的”音或“嘶哑”音中以外，还在“假声”或“翻转”的音中也出现的。像“紧张的”音、“嘶哑”音、“假声”、“翻转”那样的、以脱离平均的讲话状态(以平常发声的讲话)的讲话状态发声的声音示出，针对特定的音响特性的、远离以平均的讲话状态发声的声音的值。在有充分地含有大量且各种各样的讲话状态的声音数据的情况下，像在日本国特开 2004-279436 号公报所示的“气息性”(嘶哑)的第一共振峰周边的能量和第三共振峰周边的能量的时间性相关的例子那样，存在特定的音响特性值分布在统计上远离多数声音的分布位置的位置的情况。在特定的讲话风格或感情表现中可以观测这些分布。例如，在“气息性”的音响特性的情况下，可以确认属于表现出亲密感的声音的倾向。反而，也存在下列可能性：通过提取输入声音中的“紧张的”音、日本国特开 2004-279436 号公报所述的“气息性”(嘶哑)的音或“假声”，从而可以判断讲话者的感情、讲话态度的种类或状态。再者，也存在下列可能性：通过确定此特征性音色被检测出的部位的音韵，从而可以判断讲话者的感情或讲话态度的程度。

图 3A 以及图 3B 示出下列推定结果：针对图 3A 所示的输入“じゅっぷんほどかかります”和图 3B 所示的输入“あたたまりました”，根据推定公式来推定以“紧张的”音发声各个音拍时的“紧张容易度”的结果，所述推定公式是依据与图 1A～图 1D 相同的数据、并利用统

计学习法的一种即数量化 II 类来制作的。例如，在图 3A 的“かかります”中示出，只在概率高的音拍发生“紧张的”音，因此“愤怒”的程度小。同样，图 3B 中也示出，在“あたたま”中“紧张”的发生概率高或中等程度，因此“愤怒”的程度是小至中等程度，在“り”中“紧张”的发生概率低，因此“愤怒”的程度大。该例子中，对于学习用数据的各个音拍，将示出音拍中包含的子音以及母音的种类或像音韵的范畴那样示出音韵的类型的信息和声调句内的音拍位置作为独立变量，还将前后的音韵的信息作为独立变量。而且，将是否发生了“紧张的”音或“刺耳的声音(harsh voice)”的二值作为从属变量。该例子是下列结果：根据这些独立变量以及从属变量，依据数量化 II 类来制作推定公式，并且将发生概率分隔为低、中、高的三个阶段的情况下的结果。该例子示出，通过利用声音识别结果按每个音拍求出输入声音的特征性音色的发生概率，从而可以判断讲话者的感情或讲话态度的程度。

通过将使用依据发声时的生理性特征的特征性音色的发生概率来求出的、感情或讲话态度的种类和程度，作为感情的种类和强度的指标利用，从而可以进行因语言或地方(方言)差别或个人差别而引起的影响小的、准确的感情的判断。

下面，参照附图说明本发明的实施例。

(实施例 1)

图 4 是本发明的实施例 1 中的依据声音的感情识别装置的功能框图。图 5 是实施例 1 中的感情识别装置的工作流程图。图 6 是在特征

性音色发生指标计算规则记忆部 110 记忆有的计算规则的一个例子，图 7 是在感情种类判断基准记忆部 112 记忆有的判断基准的一个例子，图 8 是在感情强度计算规则记忆部 114 记忆有的感情强度计算规则的一个例子。

在图 4 中，感情识别装置是依据声音识别感情的装置，包括：麦克风 1、声音识别用特征量提取部 101、逆滤波器 102、周期性分析部 103、特征性音色检测部 104、特征量数据库 105、声音识别部 106、开关 107、特征性音色发生音韵确定部 108、韵律信息提取部 109、特征性音色发生指标计算规则记忆部 110、特征性音色发生指标计算部 111、感情种类判断基准记忆部 112、感情种类判断部 113、感情强度计算规则记忆部 114、感情强度计算部 115、以及显示部 116。

麦克风 1 是将输入声音转换为电信号的处理部。

声音识别用特征量提取部 101 是一种处理部，分析输入声音来提取表示谱包络的参数，例如提取 Mel 倒谱(MelCepstrum) 系数。

逆滤波器 102 是声音识别用特征量提取部 101 输出的谱包络信息的逆滤波器，并且，逆滤波器 102 是输出麦克风 1 所输入的声音的音源波形的处理部。

周期性分析部 103 是一种处理部，分析逆滤波器 102 所输出的音源波形的周期性，从而提取音源信息。

特征性音色检测部 104 是一种处理部，利用物理特性(例如，音源波形的振幅微扰或音源波形的周期微扰等)，周期性分析部 103 所输出的音源信息中检测依据讲话者的感情或讲话态度在讲话声音中出现

的、“紧张的”声音、“假声”或“气息性”(嘶哑)的声音等特征性音色。

特征量数据库 105 是一种记忆装置，保持用于声音识别的每个音韵种类的特征量，例如，保持将每个音韵的特征量的分布作为概率模型表示的数据。特征量数据库 105，由依据声音中没有特征性音色的声音数据来制作的特征量数据库、和依据声音中存在特征性音色的声音数据来制作的特征量数据库构成。例如，如下构成：依据没有特征性音色的声音数据来制作的特征量数据库叫做无特征性音色的特征量数据库 105a；依据存在“紧张的”声音的特征性音色的声音数据来制作的特征量数据库叫做具有“紧张”的特征量数据库 105b；依据存在“气息性”(嘶哑)声音的特征性音色的声音数据来制作的特征量数据库叫做具有“嘶哑”的特征量数据库 105c；依据存在“紧张的”声音的特征性音色和“气息性”(嘶哑)声音的特征性音色的两者的声音数据来制作的特征量数据库叫做具有“紧张”、“嘶哑”的特征量数据库 105d。

声音识别部 106 是一种处理部，参照特征量数据库 105，对比声音识别用特征量提取部 101 所输出的特征量和特征量数据库 105 所存储的特征量，来进行声音识别。

开关 107，按照由特征性音色检测部 104 检测出的音源波形的微扰的有无以及微扰的种类，来切换到声音识别部 106 所参照的、构成特征量数据库 105 的数据库。

特征性音色发生音韵确定部 108 是一种处理部，依据声音识别部 106 输出的音韵列信息、和特征性音色检测部 104 输出的输入声音中的特征性音色的时间位置信息，来确定在输入声音中的哪个音韵中发生

了特征性音色。

韵律信息提取部 109 是一种处理部，从逆滤波器 102 所输出的音源波形中提取声音的基频和功率。

特征性音色发生指标计算规则记忆部 110 是记忆规则的记忆部，该规则用于，将每个音韵的特征性音色的发生容易度的指标，依据该音韵的属性(例如，子音的种类、母音的种类、声调句或重读句内的位置、与声调或重读位置的关系、基频的绝对值或倾斜等)来求出。

特征性音色发生指标计算部 111 是一种处理部，依据声音识别部 106 所生成的音韵列信息、和韵律信息提取部 109 所输出的韵律信息即基频和功率，参照特征性音色发生指标计算规则记忆部 110，来计算每个输入声音的音韵的特征性音色发生指标。

感情种类判断基准记忆部 112 是记忆基准的记忆部，该基准是依据该音拍以及相邻的音拍的特征性音色的种类、和特征性音色发生指标的组合来判断感情的种类时的基准。

感情种类判断部 113 是一种处理部，根据特征性音色发生音韵确定部 108 所生成的特征性音色发生位置信息，参照感情种类判断基准记忆部 112 的基准，来判断每个音拍的感情的种类。

感情强度计算规则记忆部 114 是记忆规则的记忆部，该规则用于依据特征性音色的发生指标、和输入声音的特征性音色发生位置信息，来计算感情或讲话态度的程度。

感情强度计算部 115 是一种处理部，依据特征性音色发生音韵确定部 108 所生成的、输入声音中发生了特征性音色的音韵的信息、和

特征性音色发生指标计算部 111 所计算的每个音韵的特征性音色发生指标，参照感情强度计算规则记忆部 114，来输出感情或讲话态度的程度和感情种类、以及音韵列。

显示部 116 是一种显示装置，显示感情强度计算部 115 的输出。

按照图 5 说明如上述构成的依据声音的感情识别装置的工作。

首先，由麦克风 1 输入声音(步骤 S1001)。声音识别用特征量提取部 101，分析输入声音，来作为声音识别用的音响特征量提取 Mel 倒谱系数(步骤 S1002)。其次，逆滤波器 102，设定参数，以便作为在步骤 S1002 所生成的 Mel 倒谱系数的逆滤波器，并且，使在步骤 S1001 麦克风所输入的声音信号通过，来提取音源波形(步骤 S1003)。

周期性分析部 103，依据在步骤 S1003 所提取的音源波形的周期性，像例如日本国特开平 10-197575 号公报所记载的技术那样，依据具有低频率侧缓且低频率侧急的遮断特性的断路特性的过滤器输出的振幅调制的大小和频率调制的大小来计算基波分量，将输入声音中具有周期性的信号的时间区域作为周期性信号区间输出(步骤 S1004)。

特征性音色检测部 104，对于在步骤 S1004 由周期性分析部 103 所提取的周期性信号区间，在本实施例中检测音源波形的微扰中的音源波形的基频微扰(jitter)以及音源波形的高频域成分的微扰(步骤 S1005)。并且，例如，利用以日本国特开平 10-19757 号公报的方式求出的瞬时频率来检测基频微扰。而且，例如，像日本国特开 2004-279436 号公报所记载的技术那样，依据利用正规化振幅指数的方法来检测音源波形的高频域成分的微扰，该正规化振幅指数是以基频将下列值频

正规化的指数，该值是音源波形的峰间的振幅除以音源波形的微分的振幅的最小值(最大负峰值)的值。

依据在输入声音的周期性信号区间是否检测出音源波形的频率微扰或音源波形的高频域成分的微扰，切换开关 107 来连接特征量数据库 105 内的适当的特征量数据库和声音识别部 106(步骤 S1006)。即，在步骤 S1005 检测出音源波形的频率微扰的情况下，通过开关 107 连接特征量数据库 105 中的具有“紧张”的特征量数据库 105b 和声音识别部 106。在步骤 S1005 检测出音源波形的高频域成分的微扰即气息性(嘶哑)的成分的情况下，通过开关 107 连接特征量数据库 105 中的具有“嘶哑”的特征量数据库 105c 和声音识别部 106。在步骤 S1005 检测出音源波形的频率微扰和音源波形的高频域成分的微扰的两者的情况下，通过开关 107 连接特征量数据库 105 中的具有“紧张”、“嘶哑”的特征量数据库 105d 和声音识别部 106。而且，在步骤 S1005 未检测音源波形的频率微扰和音源波形的高频域成分的微扰的两者的情况下，通过开关 107 连接特征量数据库 105 中的无特征性音色的特征量数据库 105a 和声音识别部 106。

声音识别部 106，参照特征量数据库 105 中的通过开关 107 连接的特征量数据库，利用在步骤 S1002 提取的 Mel 倒谱系数来进行声音识别，作为识别结果一并输出音韵列和输入声音中的时间位置信息(步骤 S1007)。

特征性音色发生音韵确定部 108 依据声音识别部 106 所输出的带有时间位置信息的音韵列信息、和特征性音色检测部 104 所输出的输

入声音中的特征性音色的时间位置信息，来确定在输入声音中的哪个音韵中发生了特征性音色(步骤 S1008)。

另外，韵律信息提取部 109，分析逆滤波器 102 所输出的音源波形来提取基频和音源功率(步骤 S1009)。

特征性音色发生指标计算部 111，依据声音识别部 106 所生成的带有时间位置信息的音韵列信息、和韵律信息提取部 109 所提取的基频和音源功率，将基频模式和音源功率模式的起伏、与音韵列对照，来生成对应于音韵列的声调句分隔以及声调信息(步骤 S1010)。

再者，特征性音色发生指标计算部 111，利用在特征性音色发生指标计算规则记忆部 110 记忆有的规则，按每个音韵列的音拍计算特征性音色发生指标，该规则用于依据子音、母音、声调句中的音拍位置、从声调核的相对位置等音拍属性来求出特征性音色的发生容易度(步骤 S1011)。例如，特征性音色发生指标的计算规则是通过下列处理来制作的：依据含有带有特征性音色的声音的声音数据，将音拍属性作为解释变量、将是否发生了特征性音色的二值作为从属变量，利用处理质性数据的一种统计学习法、即利用数量化 II 类来进行统计学习，从而生成模式，该模式可以以数值表示依据音拍属性的特征性音色的发生容易度。

例如，特征性音色发生指标计算规则记忆部 110，如图 6 按每个特征性音色的种类记忆统计学习结果。特征性音色发生指标计算部 111，根据各个音拍的属性，适用在特征性音色发生指标计算规则记忆部 110 记忆有的统计模型，来计算特征性音色发生指标。在输入声音是如图

3B 所示的“あたたまりました”的情况下，特征性音色发生指标计算部 111，如下求出开头的音拍“あ”的属性的得分：由于“无子音”因此子音的得分为-0.17；由于母音是“ア”因此母音的得分为 0.754；由于在“あたたまりました”的声调句中的正顺序位置的第一个音拍，因此正顺序位置的得分为 0.267；由于在声调句中的反顺序位置的第八个音拍，因此反顺序位置的得分为 0.659。而且，特征性音色发生指标计算部 111，通过将这些得分相加，从而计算开头的音拍“あ”的特征性音色发生指标。特征性音色发生指标计算部 111，通过对各个音拍进行相同处理，从而计算各个音拍的特征性音色发生指标。据此，依据各个音拍的属性可以计算下列特征性音色发生指标：开头的“あ”为 1.51(=-0.17+0.754+0.267+0.659)；下一个“た”为 0.79；第三个音拍的“た”为 0.908。

感情种类判断部 113，依据特征性音色发生音韵确定部 108 所生成的以音韵为单位描述的特征性音色发生位置信息来确定输入声音中的特征性音色发生种类，参照如图 7 描述的感情种类判断基准记忆部 112 的信息，来确定输入声音中包含的发生了特征性音色的音拍中的感情种类(步骤 S1012)。在输入声音的“あたたまりました”中“あたたまりま”是“紧张的”声音、且在其它音拍没有以特征性音色的发声的情况下，只针对以特征性音色发声的音拍、根据图 7 的表来判断感情，从而以音拍为单位识别感情的变化。对于图 3B，在根据图 7 对“あ”进行计算的情况下，由于没有该音拍“あ”前一个音拍，因此向该音拍的“紧张”的发生指标 1.51 加上后一个音拍的“紧张”的发生指标

0.79 的一半即 0.395，从而成为 1.905。而且，在相邻的音拍不存在“紧张”的发生。据此，由于对“紧张”的计算值为正、对“紧张”的计算值为 0，因此该音拍中包含的感情被判断为“愤怒”。同样，对于第二个音拍的“た”，向该音拍的 0.79 加上前一个音拍的 1.51 一半的 0.755 和后一个音拍的 0.91 一半的 0.455，从而成为 2.0，因此与第一个音拍相同，感情被判断为“愤怒”。

然而，对于图 3A 所示的“じゅっぷんほどかかります”的输入声音，在“ほ”以“嘶哑”发声，在前一个音拍没有以特征性音色的发声，但是，在后一个“ど”以“紧张”发声。因此，对于“ほ”，该音拍的“嘶哑”的发生指标 2.26 和后一个音拍的“紧张”的发生指标 0.73 的一半的 0.365 一起判断，根据图 7 的表，对于“ほ”的部位、以及同样对于“ど”的部位，输入声音被判断为包含“欢闹、愉快而兴奋”的感情。但是，“ほど”后面的“かか”的部位，特征性音色只检测出“紧张”，根据图 7 的表被判断为包含“愤怒”的感情，可以以音拍为单位跟随用户对系统讲话中而变化的感情。

在输入声音是“あたたまりました”的情况下，针对在步骤 S1011 所计算的每个音拍的特征性音色发生指标的值(例如，开头的“あ”为 1.51、其次的“た”为 0.79、第三个音拍的“た”为 0.908)，参照如图 8 所述的感情强度计算规则记忆部 114 的感情强度计算规则时，开头的“あ”的“紧张”的发生指标为 1.51 即 0.9 以上，因此判断为“紧张”容易度“高”。如图 3B 输入声音的“あたたまりました”中以“紧张的”声音发声“あたたまりま”的情况下，由于在“紧张”容易度大

的“あ”“紧张”了，因此“愤怒”的感情强度低。对于其次的“た”，“紧张”的发生指标为 0.79，因此中等程度的“紧张”容易度、中等程度的“愤怒”；对于第三个音拍的“た”，发生指标为 0.908，因此“紧张”容易度高、“愤怒”的感情强度低。如此，按每个音拍计算感情强度(步骤 S1013)，与在步骤 S1012 进行感情判断时相比，可以求出更详细的感情强度的变化。显示部 116 显示在步骤 S1013 所计算的作为感情种类判断部 113 的输出的每个音拍的感情强度(步骤 S1014)。

对于如图 3A 的输入，在步骤 S1012 中，“じゅっぷんほどかかります”的“ほ”依据“嘶哑”的发生指标 2.26 和“紧张”的发生指标 0.365 被判断为“欢闹、愉快而兴奋”，参照如图 8 所述的感情强度计算规则记忆部 114 的规则时，将“ほ”的“紧张”的发生指标和“嘶哑”的发生指标相乘的值为 0.8249，“欢闹、愉快而兴奋”的强度弱。而且，对于“ど”，“紧张”的指标是该音拍的 0.73 和后一个音拍的 1.57 的一半相加而得的 1.515，“嘶哑”的指标是前一个音拍“ほ”的指标 2.26 的一半 1.13，将这些值相乘的值为 1.171195，因此“欢闹、愉快而兴奋”的强度弱。对于后面的“か”，紧张的指标是将前一个音拍的指标的一半、后一个音拍的指标的一半和该音拍的指标相加而得的 2.55，因此“愤怒”的强度被判断为“弱”。

在此，说明图 8 所示的在感情强度计算规则记忆部 114 记忆有的感情强度计算规则的制作方法中、指标范围和“紧张”容易度和感情强度的关系的制作方法。图 9 是有“紧张”的音拍的发生频度和没有“紧张”的音拍的发生频度和“紧张”容易度的指标的值的关系、以

及感情的强度(弱度)和指标的值的关系的模式图。在图 9 中设定，在横轴按每个音拍求出的“紧张”容易度的指标，越靠右边越容易“紧张”。而且，在纵轴示出声音中的具有“紧张”或没有“紧张”的音拍的发生频度以及每个音拍的“紧张”概率。而且，在图表中左轴示出具有“紧张”或没有“紧张”的音拍的发生频度，在图表中右轴示出每个音拍的“紧张”概率。在图表的曲线中，实线是示出依据实际声音数据制作的、指标的值和具有“紧张”的音拍的发生频度的关系的函数，细虚线是示出依据实际声音数据制作的、指标的值和没有“紧张”的音拍的发生频度的关系的函数。依据两者函数来求出具有某指标的值的音拍内以“紧张”发声的频度，并将其作为“紧张”发生概率，用粗虚线以百分率示出的“感情的弱度”。发生概率即“感情的弱度”的特性是，发生指标越小感情越强，发生指标越大感情越弱。对如图 9 所示的依据发生指标而变化的“感情的弱度”的函数，依据实际声音数据来设定感情强度的范围，依据函数来求出对应于设定了的感情强度范围的边界的发生指数，从而制作如图 8 所示的表。

并且，在图 8 所示的感情强度计算规则记忆部 114 中利用依据“感情的弱度”的函数制作的表来计算感情强度，但也可以记忆图 9 所示的函数、且依据函数来直接计算“感情的弱度”即函数强度。

根据所述结构，从所输入的声音中作为反映了感情的特征性音色提取音源微扰，保持含有特征性音色的特征量数据库和不含有特征性音色的特征量数据库，依据音源微扰的有无来切换特征量数据库，从而提高声音识别精度。另一方面，根据依据声音识别结果来求出的特

征性音色的发生容易度和实际上的输入声音的音源微扰的有无的比较结果，在容易发生特征性音色的部位实际上发生了特征性音色的情况下判断为感情的强度低，在不易发生特征性音色的部位发生了特征性音色的情况下判断为感情的强度高。据此，在不受语言差别、个人差别以及地方差别的影响的情况下，可以从输入声音中准确地识别声音的讲话者的感情的种类和强度。

而且，在利用依据无表情的声音数据作出的特征量数据库的情况下，对有感情表现的声音中出现的特征性音色的声音识别精度低，但是，通过切换到依据含有特征性音色的声音作出的特征量数据库，从而声音识别精度会提高。而且，通过识别精度的提高，利用音韵列来计算的特征性音色的发生容易度的计算精度也会提高。因此，感情强度的计算精度也会提高。再者，通过以音拍为单位检测特征性音色、以音拍为单位进行感情识别，可以以音拍为单位跟随输入声音中的感情的变化。因此，在将系统用于对话控制等的情况下，确定用户即讲话者在对话动作过程中对哪个事件怎样反应了时有效的。如此依据输入声音可以仔细地把握用户的感情的变化，因此，例如，按照用户的愤怒的强度，将系统侧的输出声音变为像“大变申し訳ございませんが・・・(实在对不起・・・)”那样的更礼貌的道歉、或像“お手数ではございますが・・・(给您添麻烦・・・)”那样的更礼貌的请求的表现，从而可以使用户的感情处于平常状态，并且可以作为对话接口进行顺利的工作。

(实施例 1 的变形例)

示出本发明的实施例 1 的变形例。图 10 是本发明的实施例 1 中的依据声音的感情识别装置的变形例的功能框图。图 11 是实施例 1 中的变形例的依据声音的感情识别装置的工作流程图。图 12 是所输入的声音的音韵列和以特征性声音发声的音拍以及其“紧张”的发生指标和“嘶哑”的发生指标的值的模式图。图 13 示出在感情强度计算规则记忆部 132 记忆有的判断感情的种类时的基准的信息的一个例子。

图 10 所示的感情识别装置具有与图 4 所示的实施例 1 涉及的感情识别装置相同的结构，但是一部分的结构不同。即，图 4 中的感情种类判断基准记忆部 112 被替换为感情种类判断规则记忆部 132。而且，感情种类判断部 113 和感情强度计算部 115 被替换为感情种类强度计算部 133。再者，结构上没有感情强度计算规则记忆部 114，而感情种类强度计算部 133 参照感情种类判断规则记忆部 132。

如上构成的依据声音的感情识别装置，在实施例 1 中的步骤 S1011 按每个音拍进行特征性音色发生指标的计算。

对于如图 12 的例子，在提取特征性音色的“紧张”和“嘶哑”、只根据其频度来判断感情的情况下，音拍数多的“紧张”大大影响到判断，因此判断为典型地出现“紧张”的“愤怒”的感情的声音，而系统进行道歉的对应。但是，实际上所输入的声音带有中等程度的“欢闹、愉快而兴奋”的感情，因此，对话系统应该提供用于用户与系统更享受会话的信息。

例如，如图 12，在 24 个音拍中以“紧张”发声的音拍存在 5 个音拍、以“嘶哑”发声的音拍存在 3 个音拍的情况下，通过与步骤 S1011

相同的方法按每个音拍算出“紧张”和“嘶哑”的特征性音色发生指数。“紧张”的特征性音色发生指数的倒数的和为4.36。另外，“嘶哑”的特征性音色发生指数的倒数的和为 4.46。此意味着，虽然对于检测出的特征性音色的音拍数“紧张”的声音多，但是，在更不易嘶哑的声音中也发生了“嘶哑”的声音，即引起“嘶哑”的感情更强。进一步，感情种类强度计算部 133，根据图 13 所示的感情种类判断规则来判断感情的种类和强度(步骤 S1313)。

而且，也可以将一个种类的特征性音色的指标平均化。例如，如图 3B，8 个音拍中以“紧张”发声的音拍存在 5 个音拍，而未发生其它特征性音色。若在发生了“紧张”和“嘶哑”的特征性音色时相同计算，则“紧张”的特征性音色发生指标的倒数(第一个音拍的“あ”0.52、第二个音拍的“た”0.50、第三个音拍的“た”0.56、第四个音拍的“ま”1.04、第五个音拍的“り”6.45、第六个音拍的“ま”1.53)的和为 10.6。依据图 13 所示的感情强度计算规则来得知，感情为“愤怒”、强度为“弱”。对于实施例 1，在图 3B 中，第五个音拍的“り”的特征性音色发生指标为-0.85，依据图 8 可以判断，感情为“愤怒”、强度为“强”。所述感情的强度的判断结果，与像实施例 1 那样按每个音拍进行判断时不同。变形例中由对话系统判断输入声音整体的感情的种类和强度，该方法在人和对话系统间的对话短且单纯的情况下有效。像实施例 1 那样，按每个音拍判断感情的种类和强度、而得到感情的种类或强度的变化，该方法在会话的内容复杂或会话长的情况下非常重要。但是，在对非常单纯的会话利用对话系统的情况下，判断

输入声音整体的感情的种类和强度的方法是有效的。例如，设想进行售票的对话系统。在此，目的为如下进行对话：对话系统打听“何枚ですか？(要几张?)”，对此用户答应“二枚お願いします。(要两张。)”。在此情况下，判断“二枚お願いします。”的输入声音整体的感情的种类和强度，在系统没有识别声音的情况下，进行如下对应：对话系统进行按照感情的种类和强度的道歉，而使用户再次答应，从而对话系统可以有效地工作。因此，本实施例的只利用一个种类的特征性音色的指标来判断输入声音整体的感情的种类和强度的声音识别系统，对短的会话或单纯的会话的对话系统等有效的。

再者，依据每个音拍的特征性音色的种类的、各个音拍的指标的倒数的和，来可以求出用于感情的判断的数值。或者，将在输入声音的特征性音色发生位置的特征性音色发生指标的值按每个特征性音色种类平均化，将输入声音的总音拍数中占有的、发生了特征性音色的音拍数作为特征性音色频度求出，将此倒数与预先求出了的特征性音色发生指标的平均值相乘，从而可以求出用于感情的判断的数值。或者，也可以是，将在输入声音的特征性音色发生位置中的特征性音色发生指标的值按每个特征性音色种类平均化，通过平均值的倒数乘以特征性音色发生频度等，从而求出用于感情的判断的数值。对于用于感情的判断的数值的求出方法，若特征性音色的发生容易度作为权重对感情的判断有效、且感情种类判断规则记忆部 132 记忆有符合计算方法的判断基准，则可以采用其它方法。

并且，在此，在步骤 S1313 求出特征性音色发生指标的强度，感

情种类判断规则记忆部 132 记忆依据每个特征性音色的强度差的判断规则，但也可以判断规则基准由特征性音色发生指标的强度比构成。

根据该结构，从所输入的声音中作为反映了感情的特征性音色提取音源微扰。另一方面，通过依据音源微扰的有无来切换特征量数据库，从而可以进行声音识别精度提高了的声音识别。利用声音识别结果可以计算特征性音色的发生容易度。在容易发生特征性音色的部位实际上发生了特征性音色的情况下判断为感情的强度低，在不易发生特征性音色的部位发生了特征性音色的情况下判断为感情的强度高，依据输入声音的一个讲话中检测出的特征性音色的发生指标，在不受个人差别或地方差别的影响的情况下，可以准确地识别以其讲话整体表示的讲话者的感情的种类和强度。

(实施例 2)

对于本发明的利用声音中的特征性音色的感情识别，通过利用声音识别结果的音韵列来求出特征性音色发生指标，从而可以进行精度高的感情识别。然而，声音识别中存在的问题是：在很多情况下，感情带有的特征性音色远离一般的音响模型，因此声音识别精度会降低。在实施例 1 中，通过具备并切换含有特征性音色的音响模型来解决了所述问题，但是，还存在的问题是：由于需要具备多种音响模型，因此数据量变大，而且，用于生成音响模型的脱机工作会增大。为了解决所述实施例 1 的问题，本实施例示出用于如下构成的结构：利用语言模型校正依据音响模型的识别结果，来提高识别精度，根据准确的声音识别结果的音韵列，来求出特征性音色发生指标，从而进行精度

高的感情识别。

图 14 是本发明的实施例 2 的依据声音的感情识别装置的功能框图。图 15 是本发明的实施例 2 的依据声音的感情识别装置的工作流程图。图 16A 至图 16C 是实施例 2 的工作的具体例子。

在图 14 中，对于与图 4 相同的部分省略说明，只对与图 4 不同的部分进行说明。在图 15 中也是，对于与图 5 相同的部分省略说明，只对与图 5 不同的部分进行说明。

在图 14 中，感情识别装置在结构上，与图 4 的功能框图中相比：没有韵律信息提取部 109 以及开关 107；特征量数据库 105 被替换为音响特征量数据库 205；追加了语言特征量数据库 206；声音识别部 106 被替换为连续单词声音识别部 207，该连续单词声音识别部 207 依据音响特征量和语言模型的语言特征量来，不仅对音韵进行识别、而对语言信息也进行识别，上述以外与图 4 相同。

根据图 15 说明如此构成的依据声音的感情识别装置的工作。对于与图 5 相同的工作省略说明，只对不同的部分进行说明。

由麦克风 1 输入声音(步骤 S1001)，声音识别用特征量提取部 101 提取 Mel 倒谱系数(步骤 S1002)。逆滤波器 102 提取音源波形(步骤 S1003)，周期性分析部 103 将输入声音中具有周期性的信号的时间区域作为周期性信号区间输出(步骤 S1004)。特征性音色检测部 104，对周期性信号区间检测音源波形的微扰，例如检测音源波形的基频微扰(jitter)以及音源波形的高频域成分的微扰(步骤 S1005)。连续单词声音识别部 207，参照记忆音响模型的音响特征量数据库 205 和记忆语言模

型的语言特征量数据库206，利用在步骤S1002提取的Mel倒谱系数，从而进行声音识别。例如，连续单词声音识别部207，依据利用音响模型和语言模型的、利用概率模型的声音识别方法，从而进行声音识别。

一般而言，

(公式1)

$$\hat{W} = \arg\max_{w} P(Y/W)P(W)$$

W：所指定的单词系列

Y：音响上的观测值系列

P(Y/W)：依据单词系列被附加条件的、音响上的观测值系列的概率(音响模型)

P(W)：对所假设的单词系列的概率(语言模型)

如此，选择音响模型和语言模型的乘积最高的单词系列来进行识别。若取对数，

(公式2)

$$\hat{W} = \arg\max_{w} \log P(Y/W) + \log P(W)$$

则可以将(公式1)表示为(公式2)。

音响模型和语言模型的平衡不一定相等，因此需要附加对两者模型的权重。一般而言，作为两者权重的比率设定语言模型的权重，

(公式3)

$$\hat{W} = \arg\max_{w} \log P(Y/W) + \alpha \log P(W)$$

α：音响模型和语言模型的的两者模型中的语言模型的权重

如此表示。在一般的识别处理中，语言模型的权重α在时间上具有一定的值。然而，连续单词声音识别部207获得在步骤S1005检测出的特征性音色的发生位置的信息，根据像(公式4)那样所示的、按每个单词变更语言模型权重α，

(公式4)

$$\hat{W} = \arg\max_{w} \log P(Y/W) + \sum_{i=1}^{n} \alpha_i \log P(w_i \mid w_1 \cdots w_{i-1})$$

第w_i：第i个单词

第α_i：适用于第i个单词的语言模型的权重

根据如此模型来进行连续单词声音识别。在参照音响特征量数据库和语言特征量数据库来进行声音识别时，在进行声音识别的帧包含特征性音色的情况下，将语言模型的权重α变大、将音响模型的权重相对地变小(步骤 S2006)，从而进行声音识别(步骤 S2007)。通过将语言模型的权重变大、将音响模型的权重变小，从而可以减少如下影响：因在特征性音色的发生位置不符合音响模型而引起的识别精度降低。连续单词声音识别部207，对于对输入声音进行了声音识别的结果即单词列以及音韵列，根据单词的读法信息、声调信息以及词类信息来推测声调句边界和声调位置(步骤S2010)。

例如，如图16A所示，输入声音的音韵列是“なまえをかくえんぴつがほしいんです”，在输入其中“えんぴつが”的部位以特征性音色的“紧张”被发声的声音的情况下，连续单词声音识别部207获得在步骤S1005检测出的特征性音色的发生位置的信息，对于不包含特征性音色的“なまえをかく”和“ほしいんです”的部位，适用依据

不包含特征性音色的学习用数据来决定的语言模型的权重α=0.9。此时，如图 16B 所示，依据以往的连续声音识别的方法，将语言模型的权重α为一定，对以特征性音色发声的部位也适用语言模型的权重α=0.9，该权重α=0.9 是在不以特征性音色发声的情况下适用的。在以“紧张”发声的“えんぴつが”的部位作为没有“紧张”的音响模型与“えんとつ”和配合较佳的情况下，

(公式 5)

$$P(\text{えんとつ}\ |\cdots\text{書く}) < P(\text{えんぴつ}\ |\cdots\text{書く})$$

如此，作为语言模型，从句子的开头到“書く”为止的单词列后接着“えんとつ”的概率比接着“えんぴつ”的概率大。

因此，虽然为

(公式 6)

$$P(W_1) < P(W_2)$$

W_1 = 名前 を 書く えん とつ が 欲しい ん です

W_2 = 名前 を 書く えん ぴつ が 欲しい ん です

但是，语言模型的权重小，与此相对音响模型的值大，公式 3 的值为

(公式 7)

$$\log P(Y/W_1) + 0.9 \times \log P(W_1) > \log P(Y/W_2) + 0.9 \times \log P(W_2)$$

而作为识别结果“名前を書く煙突が欲しいんです”被采用。

然而，本实施例中，在利用依据不包含特征性音色的学习数据来制作的音响模型对包含特征性音色的输入声音进行识别的情况下，识别精度会降低，对此，在步骤 S2006 由连续单词声音识别部 207，通过将以“紧张”发声“えんぴつが”的部位的语言模型的权重变大，从

而进行对应。即，如图 16C 所示，通过适用依据包含“紧张”的发声的数据来制作的语言模型的权重 α =2.3，为

(公式 8)

$$\log P(Y/W_1)+\sum_{i=1}^{n}\alpha_i \log P(w_{1,i} \mid w_{1,1}\cdots w_{1,i-1}) < \log P(Y/W_2)+\sum_{i=1}^{n}\alpha_i \log P(w_{2,i} \mid w_{2,1}\cdots w_{2,i-1})$$

而作为识别结果“名前を書く鉛筆が欲しいんです”被采用，从而可以获得准确的识别结果。

特征性音色发生指标计算部 111 获得连续单词声音识别部 207 所输出的音韵列、以音韵为单位描述的特征性音色发生位置、以及音韵列的声调句边界和声调位置的信息。特征性音色发生指标计算部 111，利用所获得的信息和在特征性音色发生指标计算规则记忆部 110 记忆有的规则，按每个音韵列的音拍计算特征性音色发生指标，所述规则用于依据子音、母音、声调句中的音拍位置、从声调核的相对位置等音拍属性来求出特征性音色的发生容易度(步骤 S1011)。感情种类判断部 113，依据特征性音色发生音韵确定部 208 所生成的以音韵为单位描述的特征性音色发生位置，来确定输入声音中的特征性音色发生种类，参照感情种类判断基准记忆部 112 的信息，来确定与输入声音中包含的特征性音色的种类相对应的感情种类(步骤 S1012)。感情强度计算部 115，比较以音韵为单位描述的输入声音的特征性音色发生位置和在步骤 S1011 由特征性音色发生指标计算部 111 计算的每个音拍的特征性音色发生指标，依据各个音拍的指标的大小和输入声音对应的音拍的状态的关系，根据在感情强度计算规则记忆部 114 记忆有的规则来计算每个音拍的的感情强度(步骤 S1013)。显示部 116 显示在步骤 S1013

计算的、由感情种类判断部 113 输出的每个音拍的的感情强度(步骤S1014)。

并且，在本实施例 2 中适用于不包含特征性音色的帧的语言模型的权重是 0.9、适用于以“紧张”发声的帧的语言模型的权重是 2.3，但是，若在包含特征性音色的帧中语言模型的权重相对地变大，则可以是其它的值。而且，可以对“紧张”以外的“嘶哑”“假声”等特征性音色分别设定所适用的语言模型的权重，也可以设定适用于包含特征性音色的帧的语言模型的权重、和适用于不包含特征性音色的帧的语言模型的权重的两种权重。

并且，对于实施例 2，也可以实施如实施例 1 所述的变形例。

根据这些结构，从所输入的声音中作为反映了感情的特征性音色提取音源微扰，另一方面，在存在音源微扰的情况下，考虑到不易符合音响特征量数据库内的音响模型，来将语言模型的权重系数 α 变大、而将音响模型的权重相对地变轻。据此，可以防止因音响模型不符合而引起的音韵级的识别错误，也可以提高句子程度的声音识别精度。另外，依据音源微扰的有无来判断输入声音的感情的种类，还利用声音识别结果来计算特征性音色的发声的容易度，在容易发生特征性音色的部位实际上发生了特征性音色的情况下，判断为感情的强度低，在不易发生特征性音色的部位发生了特征性音色的情况下，判断为感情的强度高。据此，在不受个人差别或地方差别的影响的情况下，可以从输入声音中准确地识别声音的讲话者的感情的种类和强度。

进一步，决定现有的语言模型和音响模型的平衡取决于语言模型

的权重。据此，与生成含有特征性音色的音响模型的情况相比，可以以少量的数据来生成特征量数据库。而且，在利用依据无表情的声音数据作出的特征量数据库的情况下，对有感情表现的声音中出现的特征性音色的声音识别精度低，但是，对于发生了特征性音色的部位，有可能音响模型不适当，因此将音响模型的权重变轻、而语言模型的权重变大。据此，将因适用不适当的音响模型而引起的影响变小，从而声音识别精度会提高。通过识别精度提高，利用音韵列来计算的特征性音色的发生容易度的计算精度也会提高。因此，感情强度的计算精度也会提高。再者，通过以音拍为单位检测特征性音色、以音韵为单位进行感情识别，从而可以以音韵为单位跟随输入声音中的感情的变化。因此，在用于对话控制等的情况下，确定用户即讲话者在对话动作过程中对哪个事件怎样反应了时有效的。

(实施方式 3)

图 17 是本发明的实施例 3 的依据声音的感情识别装置的功能框图。图 18 是本发明的实施例 3 的依据声音的感情识别装置的工作流程图。图 19 是示出本发明的实施例 3 的音韵输入方法的一个例子的图。

在图 17 中，对于与图 4 相同的部分省略说明，只对与图 4 不同的部分进行说明。在图 18 中也是，对于与图 5 相同的部分省略说明，只对与图 5 不同的部分进行说明。

在图 17 所示的感情识别装置中，图 4 中的声音识别用特征量提取部 101 被替换为特征量分析部 301。而且，没有特征量数据库 105 和开关 107，声音识别部 106 被替换为音韵输入部 306，其它结构与图 4 相

同。

在图 17 中的感情识别装置是依据声音来识别感情的装置，包括：麦克风 1、特征量分析部 301、逆滤波器 102、周期性分析部 103、特征性音色检测部 104、音韵输入部 306、特征性音色发生音韵确定部 108、韵律信息提取部 109、特征性音色发生指标计算规则记忆部 110、特征性音色发生指标计算部 111、感情种类判断基准记忆部 112、感情种类判断部 113、感情强度计算规则记忆部 114、感情强度计算部 115、表示部 116。

特征量分析部 301 是一种处理部，分析输入声音来提取表示谱包络的参数，例如提取 Mel 倒谱系数。

音韵输入部 306 是一种输入单元，用户针对输入波形的特定的区间输入对应的音韵种类，例如是鼠标或笔输入板等指点器。用户，例如，以下列方法来输入音韵种类：对在画面上出示的输入声音的波形或光谱图，利用指点器进行区间指定，由键盘输入与此区间相对应的音韵种类；或利用指向器从所显示的音韵种类的列表中选择。

根据图 5 说明如上构成的依据声音的感情识别装置的工作。

首先，由麦克风 1 输入声音(步骤 S1001)。特征量分析部 301，分析输入声音，来提取作为表示谱信息的音响特征量的 Mel 倒谱系数(步骤 S3001)。其次，逆滤波器 102，设定参数，以便作为在步骤 S1002 所生成的 Mel 倒谱系数的逆滤波器，并且，使在步骤 S1001 中麦克风所输入的声音信号通过，来提取音源波形(步骤 S1003)。

周期性分析部 103，计算在步骤 S1003 所提取的音源波形的基波分

量，依据基波分量将输入声音中具有周期性的信号的时间区域作为周期性信号区间输出(步骤 S1004)。

特征性音色检测部 104，对于在步骤 S1004 周期性分析部 103 所提取的周期性信号区间，检测音源波形的微扰(步骤 S1005)。

另外，通过音韵输入部 306，用户输入与输入声音的特定区间相对应的音韵种类(步骤 S3002)。音韵输入部 306，将与所输入的输入声音的区间相对应的音韵种类，作为输入声音的时间位置、和与此时间位置相对应的音韵信息，输出到特征性音色发生音韵确定部 108。

特征性音色发生音韵确定部 108 依据音韵输入部 306 所输出的带有时间位置信息的音韵列信息、和特征性音色检测部 104 所输出的输入声音中的特征性音色的时间位置信息，来确定在输入声音中的哪个音韵中发生特征性音色(步骤 S1008)。

另外，韵律信息提取部 109，分析逆滤波器 102 所输出的音源波形来提取基频和音源功率(步骤 S1009)。

特征性音色发生指标计算部 111，依据在步骤 S3002 所输入的带有时间位置信息的音韵列信息、和韵律信息提取部 109 所提取的基频和音源功率，将基频模式和音源功率模式的起伏、与音韵列对照，来生成对应于音韵列的声调句分隔以及声调信息(步骤 S1010)。

再者，特征性音色发生指标计算部 111，利用在特征性音色发生指标计算规则记忆部 110 记忆有的规则，按照每个音韵列的音拍来计算特征性音色发生指标，该规则用于依据子音、母音、声调句中的音拍位置、从声调核的相对位置等音拍属性来求出特征性音色的发生容易

度(步骤 S1011)。

感情种类判断部 113，依据特征性音色发生音韵确定部 108 所生成的以音韵为单位描述的特征性音色发生位置信息来确定输入声音中的特征性音色发生种类，参照感情种类判断基准记忆部 112 的信息，来确定输入声音中包含的发生了特征性音色的音拍中的感情种类(步骤 S1012)。

感情强度计算部 115，参照感情强度计算规则记忆部 114 所存储的规则来计算每个音拍的的感情强度(步骤 S1013)。与在步骤 S1012 进行感情判断时相比，可以求出更详细的感情强度的变化。显示部 116 显示在步骤 S1013 计算的、由感情种类判断部 113 输出的每个音拍的的感情强度(步骤 S1014)。

并且，在本实施例 3 中，在步骤 S1012 根据在感情种类判断基准记忆部 112 记忆有的感情种类判断基准来确定各个音韵中的感情种类，然后，在步骤 S1013 根据感情强度计算规则记忆部 114 所存储的规则来计算每个音韵的感情强度，但也可以如实施例 1 的变形例那样，计算每个音韵的特征性音色发生指标，根据其结果来计算讲话整体的感情种类和强度。

根据这些结构，从所输入的声音中作为反映了感情的特征性音色提取音源微扰，另一方面，与输入声音的特定的区间相对应的音韵种类被输入。根据由音韵列和音律信息求出的特征性音色的发的生容易度、和实际上的输入声音的音源微扰的有无，在容易发生特征性音色的部位实际上发生了特征性音色的情况下，判断为感情的强度低，在

特征性音色不易发生的部位发生了特征性音色的情况下，判断为感情的强度高。据此，在不受语言差别、个人差别或地方差别的影响的情况下，可以从输入声音中准确地识别声音的讲话者的感情的种类和强度。

并且，可以确认下列内容：韵律信息完全相同时，在将以由特征性音色的发生指标变大的趋向较大的音韵形成的特征性音色发声的声音(例如，由タ行、カ行、ダ行的ア段、エ段、オ段的音韵形成的、容易“紧张”的声音)、和以由特征性音色的发生指标变小的趋向较大的音韵形成的特征性音色发声的声音(例如，由ハ行和サ行的イ段和ウ段的音韵形成的声音)输入到本发明的感情识别装置的情况下，通过比较各个感情种类和强度的判断结果，从而算出音韵种类和韵律信息作为参数利用的特征性音色发生指标，根据特征性音色发生指标推测感情种类和强度。并且，可以确认下列内容：将使以特征性音色发声的声音连续的声音的声调位置按每一个音韵移动了的声音输入到本发明的感情识别装置的情况下，通过确认依据声调位置的移动的感情强度的变化，从而算出将音韵种类和韵律信息作为参数利用的特征性音色发生指标，根据特征性音色发生指标推测感情种类和强度。

并且，在实施例 1 以及其变形例、实施例 2、实施例 3 中，依据声音的感情识别装置，获得输入声音整体后进行处理，但也可以对麦克风 1 所输入的声音逐次进行处理。此时，在实施例 1 以及其变形例中，以作为声音识别的处理单位的音韵为逐次处理的单位，在实施例 2 中，以可语言处理的句节或句子等的单位为逐次处理的单位。

并且，在实施例 1 以及其变形例、实施例 2、实施例 3 中，依据 Mel 倒谱的逆滤波器来求出音源波形，但是，对于音源波形的求出方法，也可以利用依据 Mel 倒谱的逆滤波器的方法，即，根据声音道传达特性来求出声音道模型、并依据其逆滤波器来求出音源波形的方法，或根据音源波形的模型来求出的方法等。

并且，在实施例 1 以及其变形例、实施例 2、实施例 3 中，对于声音识别的音响特性模型利用 Mel 倒谱的参数，但也可以与此以外的声音识别方式。此时，既然可以利用 Mel 倒谱的逆滤波器来求出音源波形，也可以以其它方法来求出音源波形。

并且，在实施例 1 以及其变形例、实施例 2、实施例 3 中，将音源的频率微扰和音源的高频域微扰作为“紧张”和“嘶哑”来检测特征性音色，但也可以检测 在“粕谷英樹、楊長盛，“音源から見た声質”，日本音響学会誌 51 卷 11 号（1995），pp869-875”中列举的假声或紧张的声音等音源的振幅等“紧张”和“嘶哑”以外的特征性音色。

并且，在实施例 1 以及其变形例、实施例 2、实施例 3 中，在步骤 S1009 即特征性音色发生指标计算部 111 决定声调句边界和声调位置紧前进行基频和音源功率的提取，但是，若在步骤 S1003 中逆滤波器 102 生成音源波形后、且在步骤 S1010 中特征性音色发生指标计算部 111 决定声调句边界和声调位置之前，则可以在任何时机提取基频和音源功率。

并且，在实施例 1 以及其变形例、实施例 2、实施例 3 中，特征性音色发生指标计算部 111 作为统计学习法利用数量化 II 类，对于解释

变量，利用子音、母音、声调句中的位置、从声调核的相对位置，但是，统计学习法可以是与此以外的方法，对于解释变量，除了所述属性以外，还利用基频或功率和其模式音韵的时间长度等的连续量来计算特征性音色发生指标。

并且，在实施例 1 以及其变形例、实施例 2、实施例 3 中，对于输入声音，由麦克风 1 输入，但也可以是预先被录音、记录的声音、或从装置外部输入的声音信号。

并且，在实施例 1 以及其变形例、实施例 2、实施例 3 中，将识别了的感情的种类和强度在显示部 116 显示，但也可以将识别了的感情的种类和强度记录到记录装置、或输出到装置的外部。

本发明的依据声音的感情识别装置，通过检测依据发声器官紧张和松驰、感情、表情或讲话风格来声音的每一个细节中出现的特征性音色，从而识别输入声音的讲话者的感情或态度，所述感情识别装置有用于机器人等声音、对话接口等。而且，可以应用于呼叫中心、电话交换的自动电话对应系统等用途。再者，在声音通信时随着声音的音调而角色图像的动作变化的移动终端的应用程序中，可以应用于装载随着在声音中出现的感情的变化使角色图像的动作或表情变化的应用程序的移动终端等。

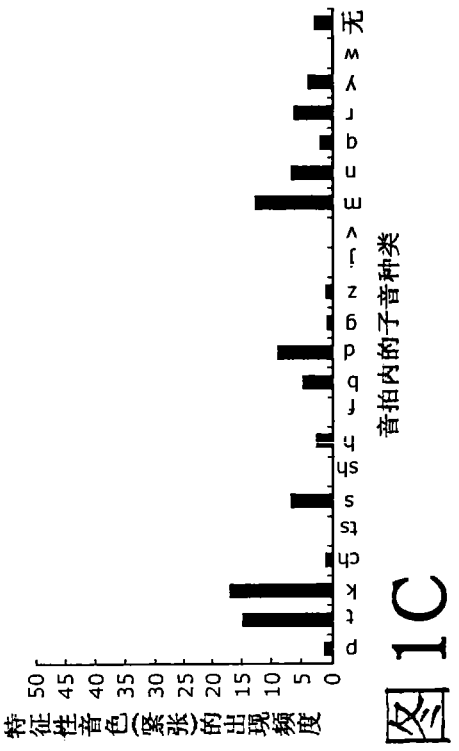


图 1C

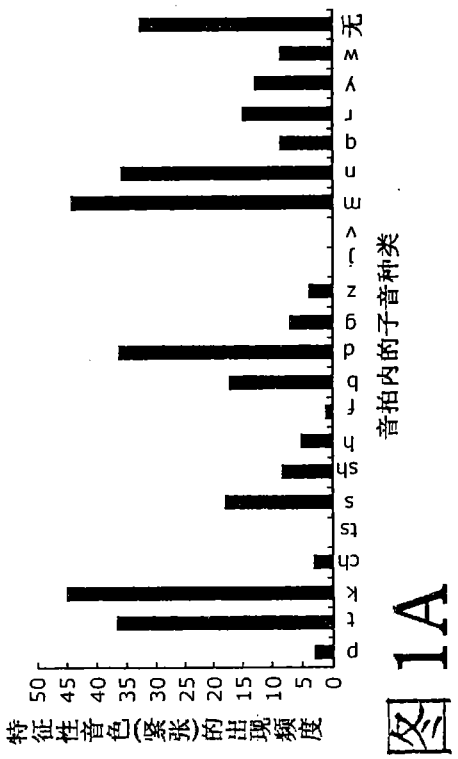


图 1A

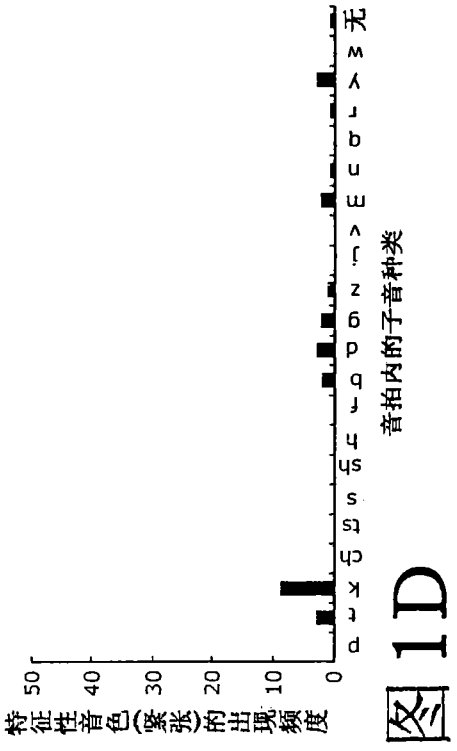


图 1D

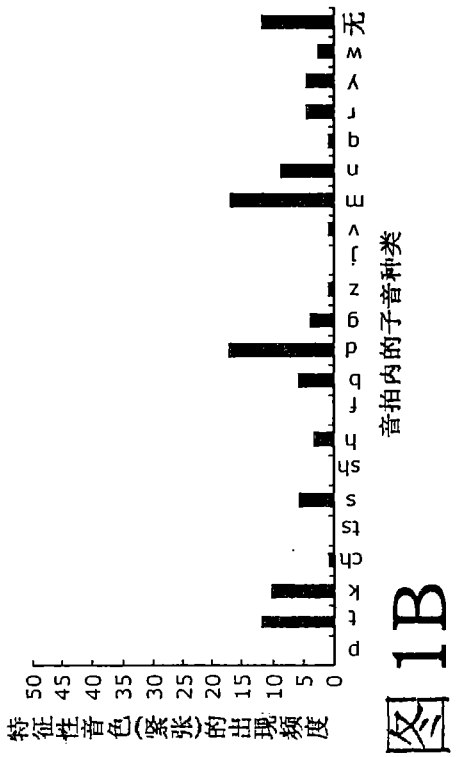


图 1B

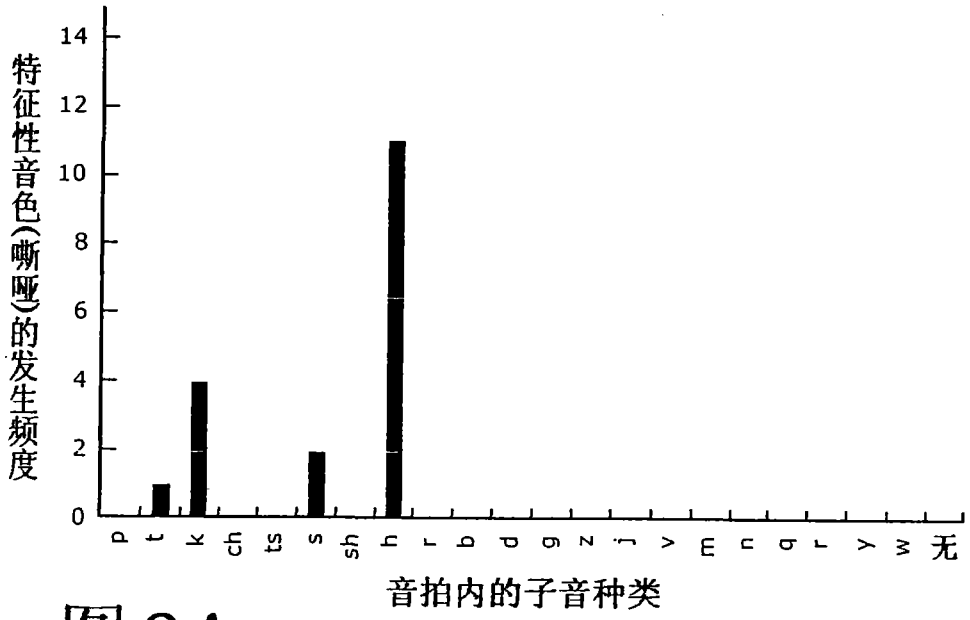

特征性音色(嘶哑)的发生频度
14
12
10
8
6
4
2
0
p t k ch ts s sh h r b d g z j v ɛ ɔ q ɾ y w 无
音拍内的子音种类

图2A

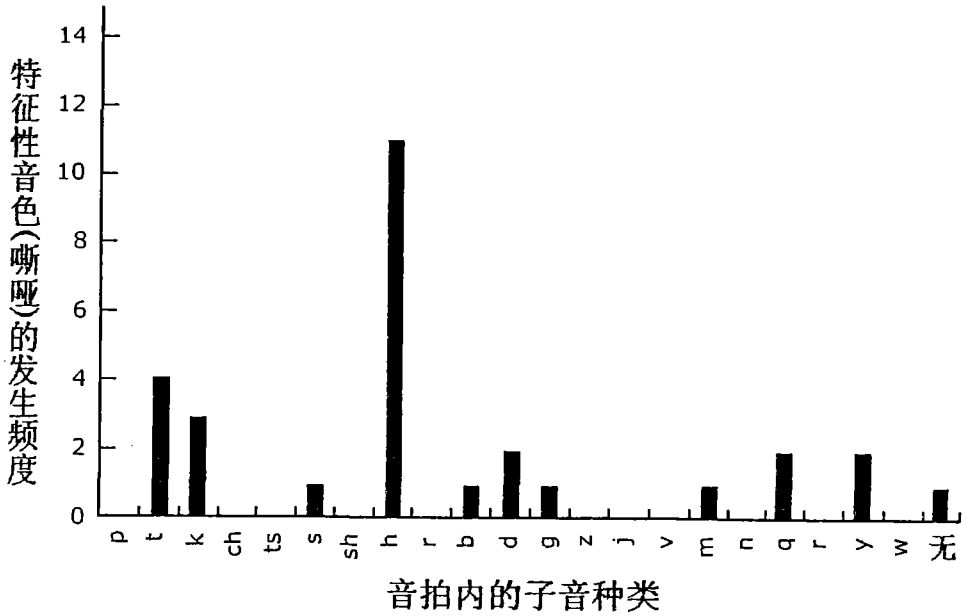

特征性音色(嘶哑)的发生频度
14
12
10
8
6
4
2
0
p t k ch ts s sh h r b d g z j v ɛ ɔ q ɾ y w 无
音拍内的子音种类

图2B

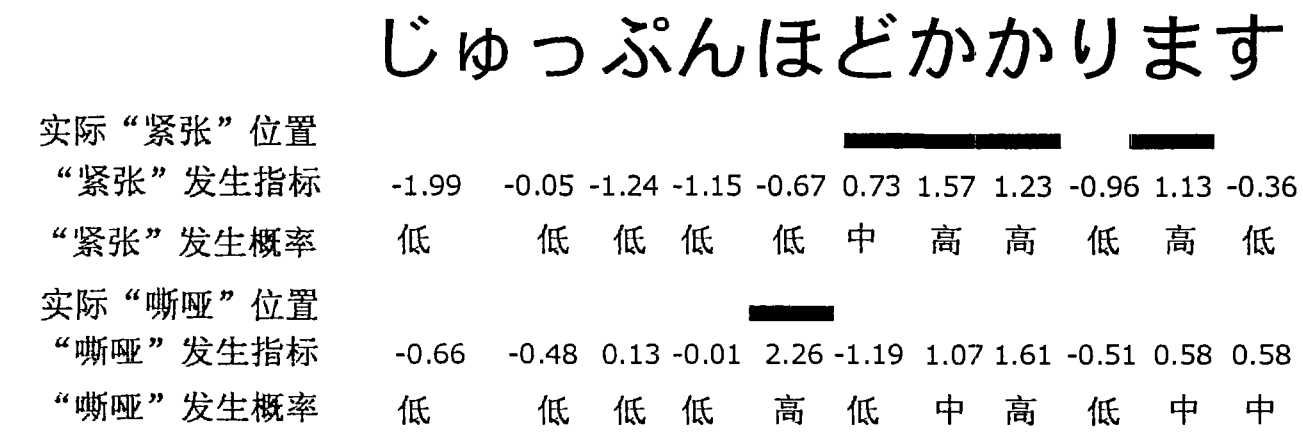

じゅっぷんほどかかります
实际“紧张”位置
“紧张”发生指标 -1.99 -0.05 -1.24 -1.15 -0.67 0.73 1.57 1.23 -0.96 1.13 -0.36
“紧张”发生概率 低 低 低 低 低 中 高 高 低 高 低
实际“嘶哑”位置
“嘶哑”发生指标 -0.66 -0.48 0.13 -0.01 2.26 -1.19 1.07 1.61 -0.51 0.58 0.58
“嘶哑”发生概率 低 低 低 低 高 低 中 高 低 中 中

图3A

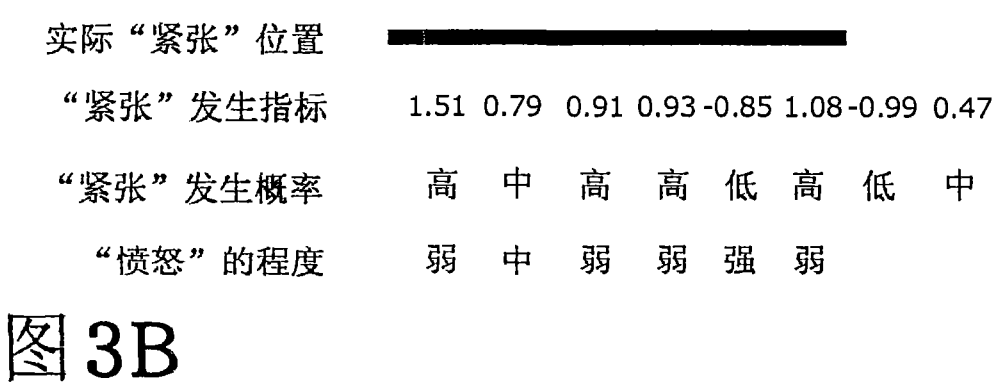

あたたまりました
实际“紧张”位置
“紧张”发生指标 1.51 0.79 0.91 0.93 -0.85 1.08 -0.99 0.47
“紧张”发生概率 高 中 高 高 低 高 低 中
“愤怒”的程度 弱 中 弱 弱 强 弱

图3B

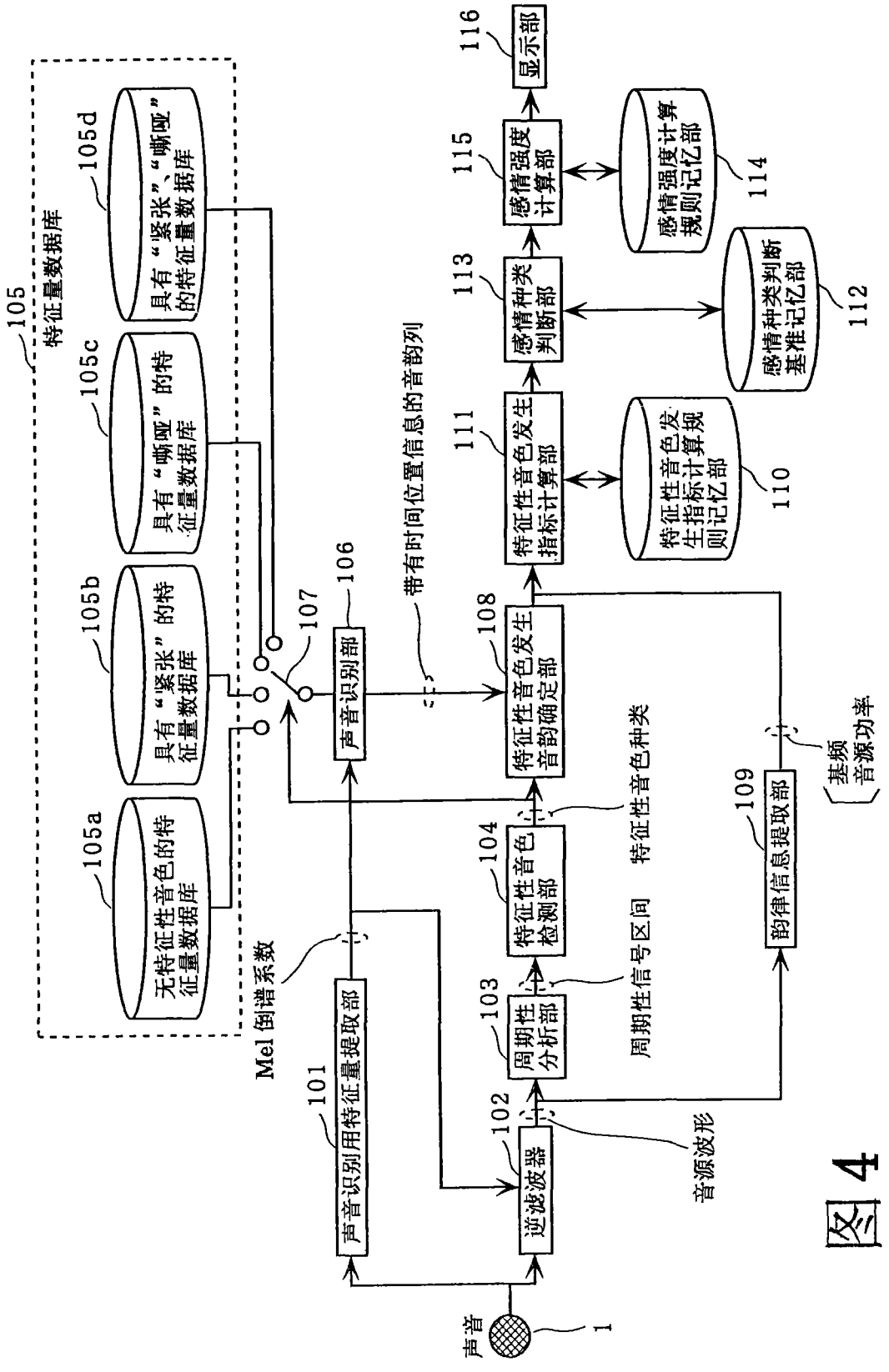

声音
1
逆波波器
102
音源波形
声音识别用特征量提取部
101
Mel 倒谱系数
周期性分析部
103
周期性信号区间
特征性音色检测部
104
特征性音色种类
韵律信息提取部
109
基频
音源功率
特征性音色发生音韵确定部
108
带有时间位置信息的音韵列
声音识别部
106
107
105
特征量数据库
无特征性音色的特征量数据库
105a
具有“紧张”的特征量数据库
105b
具有“嘶哑”的特征量数据库
105c
具有“紧张”、“嘶哑”的特征量数据库
105d
特征性音色发生指标计算部
111
特征性音色发生指标计算规则记忆部
110
感情种类判断部
113
感情种类判断基准记忆部
112
感情强度计算部
115
感情强度计算规则记忆部
114
显示部
116

图 4

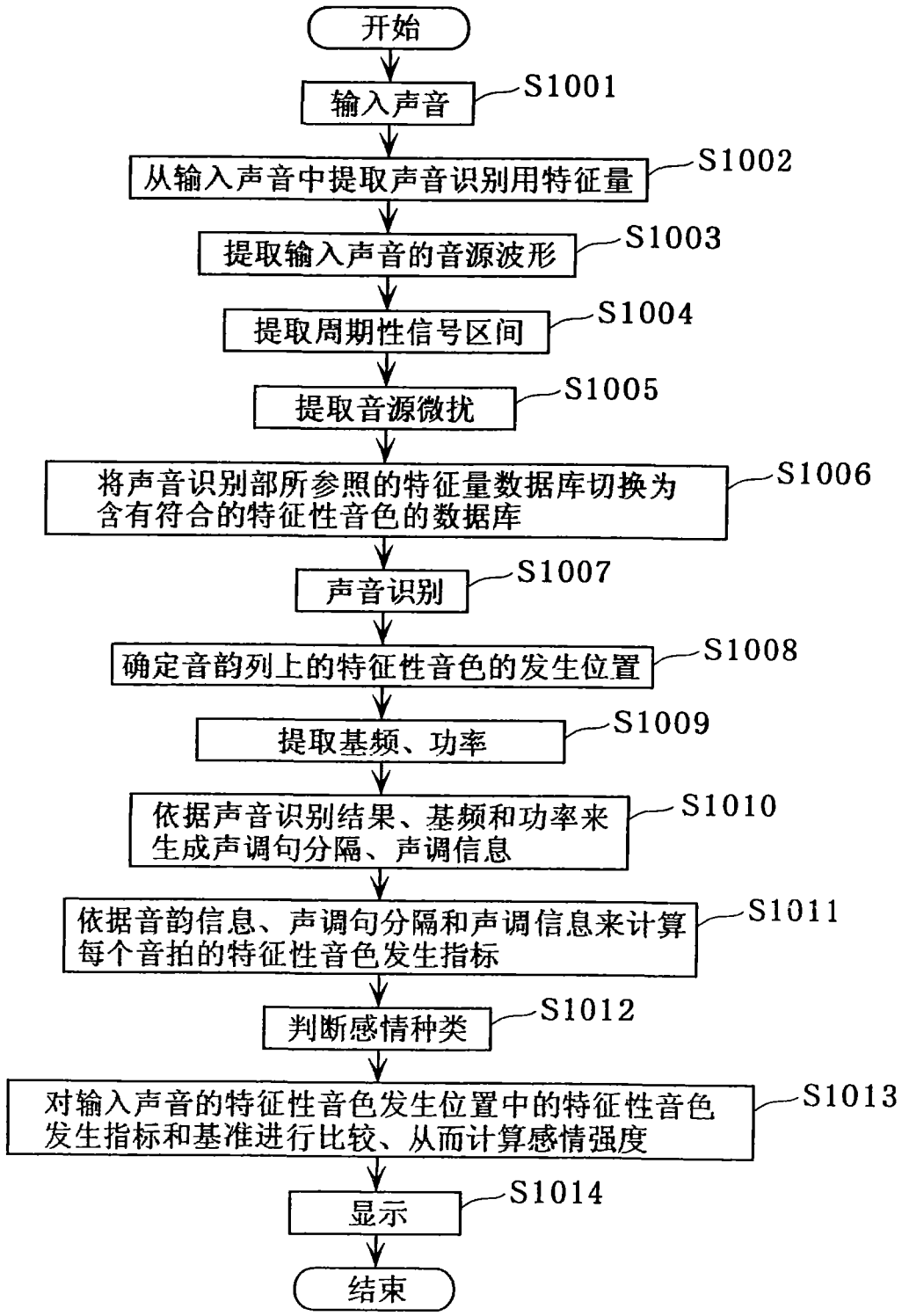

开始
输入声音
S1001
从输入声音中提取声音识别用特征量
S1002
提取输入声音的音源波形
S1003
提取周期性信号区间
S1004
提取音源微扰
S1005
将声音识别部所参照的特征量数据库切换为含有符合的特征性音色的数据库
S1006
声音识别
S1007
确定音韵列上的特征性音色的发生位置
S1008
提取基频、功率
S1009
依据声音识别结果、基频和功率来生成声调句分隔、声调信息
S1010
依据音韵信息、声调句分隔和声调信息来计算每个音拍的特征性音色发生指标
S1011
判断感情种类
S1012
对输入声音的特征性音色发生位置中的特征性音色发生指标和基准进行比较、从而计算感情强度
S1013
显示
S1014
结束

图 5

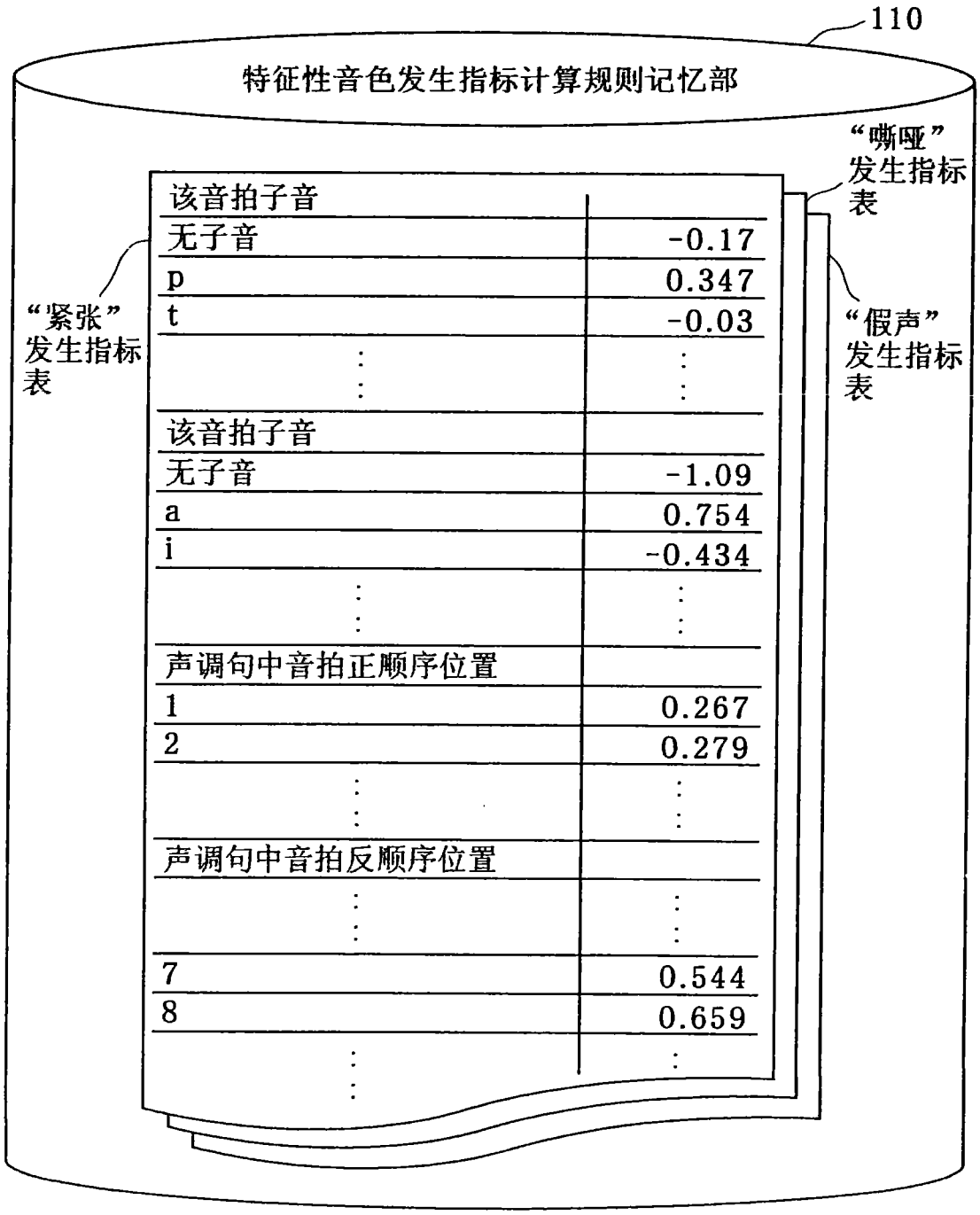

110
特征性音色发生指标计算规则记忆部
“嘶哑”发生指标表
“紧张”发生指标表
“假声”发生指标表
该音拍子音
无子音 -0.17
p 0.347
t -0.03
该音拍子音
无子音 -1.09
a 0.754
i -0.434
声调句中音拍正顺序位置
1 0.267
2 0.279
声调句中音拍反顺序位置
7 0.544
8 0.659

图6

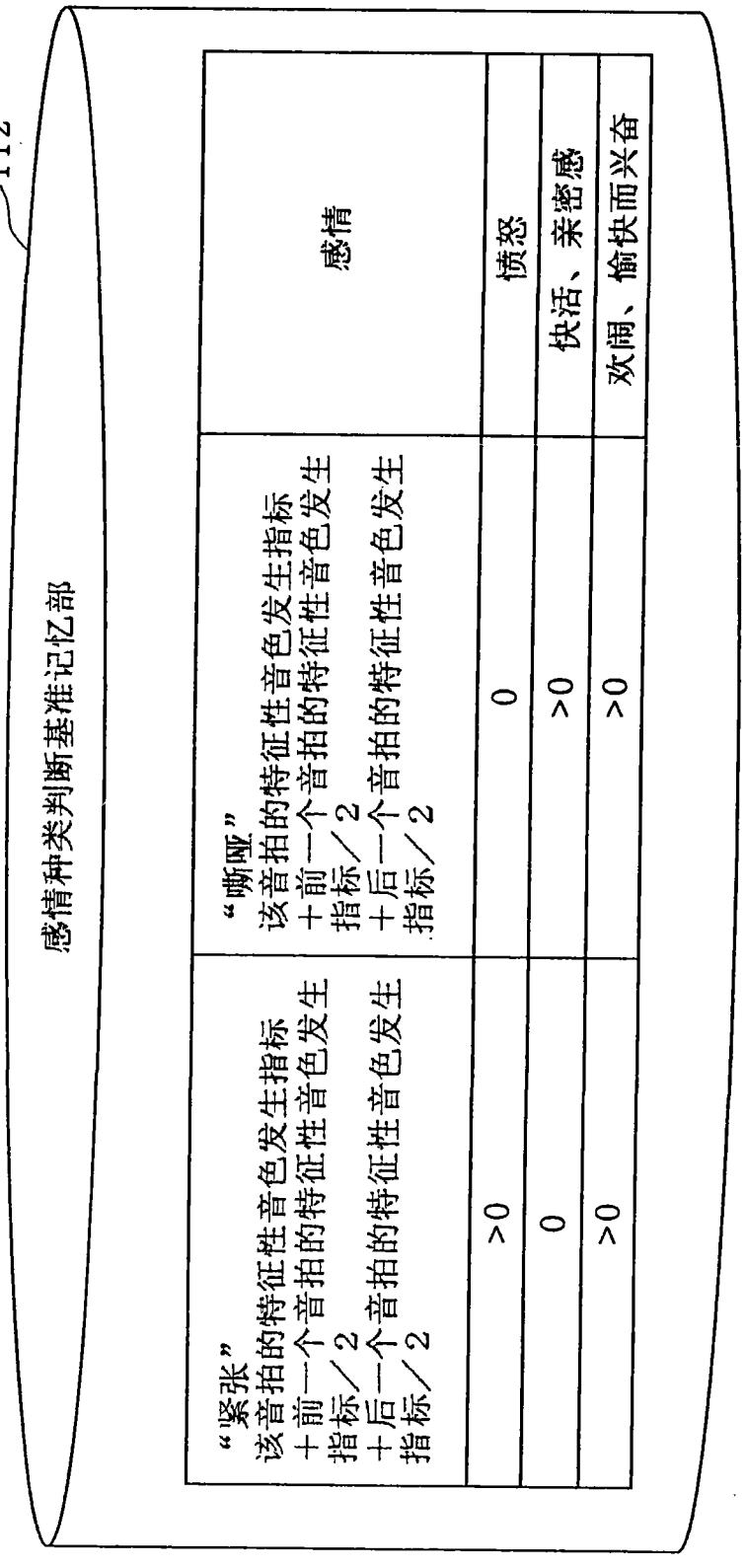

112
感情种类判断基准记忆部
“紧张”
该音拍的特征性音色发生指标
+前一个音拍的特征性音色发生
指标／2
+后一个音拍的特征性音色发生
指标／2
“嘶哑”
该音拍的特征性音色发生指标
+前一个音拍的特征性音色发生
指标／2
+后一个音拍的特征性音色发生
指标／2
感情
>0
0
愤怒
0
>0
快活、亲密感
>0
>0
欢闹、愉快而兴奋

图7

114
感情强度计算规则记忆部
只有“紧张”
指标范围
0.5>
0.5≦、0.9>
0.9≦
低
中
高
只有“嘶哑”
指标范围
0.4>
低
“紧张”和“嘶哑”
指标范围
“紧张” ×“嘶哑”
0.1>
0.1≦、0.5>
0.5≦
低
中
高
“紧张”容易度
“嘶哑”容易度
“紧张”和“嘶哑”
并存的容易度
输入声音的“愤怒”的感情强度
强
中
弱
输入声音的“快活、亲密感”的感情强度
强
输入声音的“欢闹、愉快而兴奋”
的感情强度
强
中
弱

图8

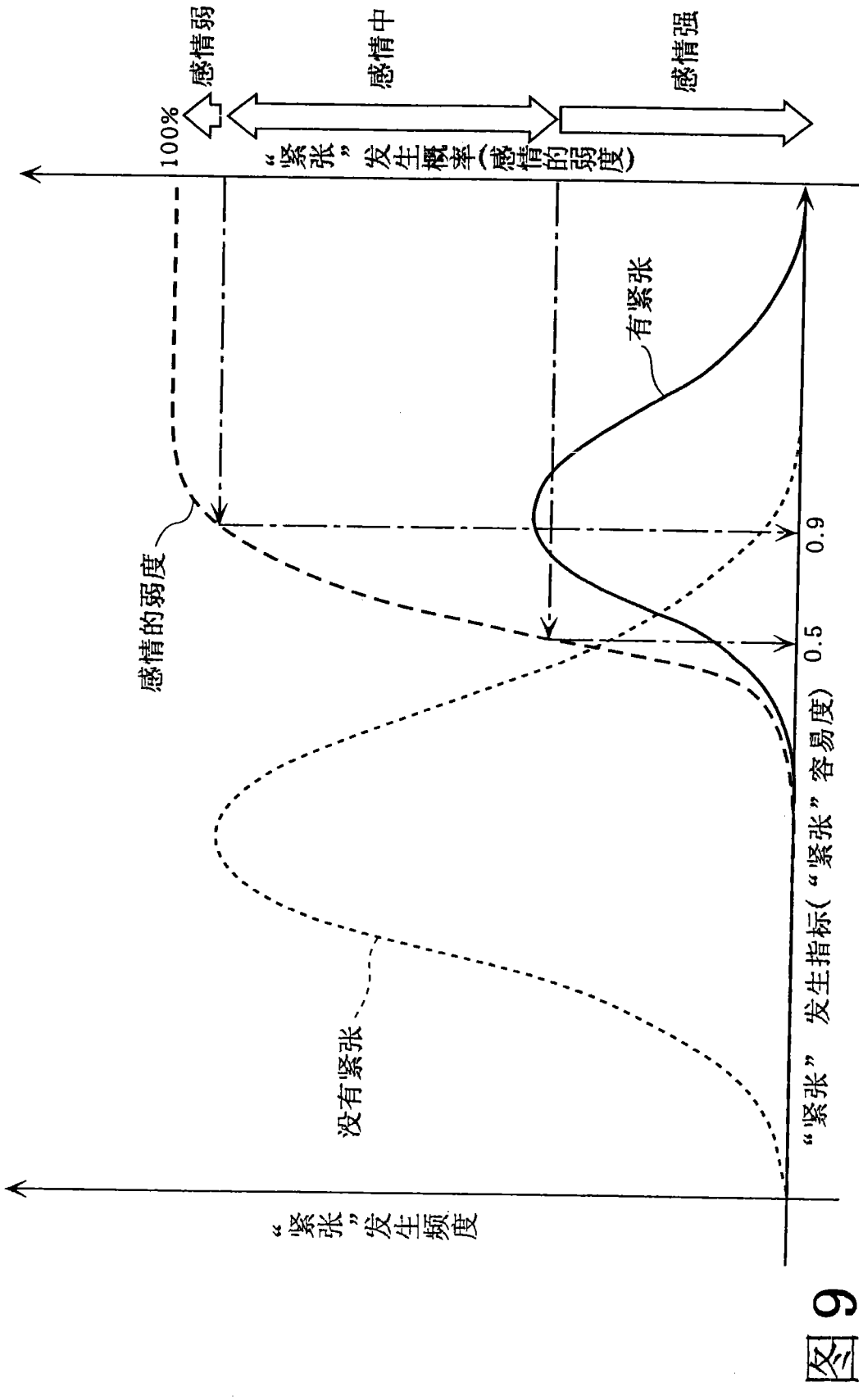

感情弱
感情中
感情强
100%
“紧张”发生概率(感情的弱度)
有紧张
感情的弱度
0.9
0.5
“紧张” 发生指标(“紧张”容易度)
没有紧张
“紧张”发生频度

图9

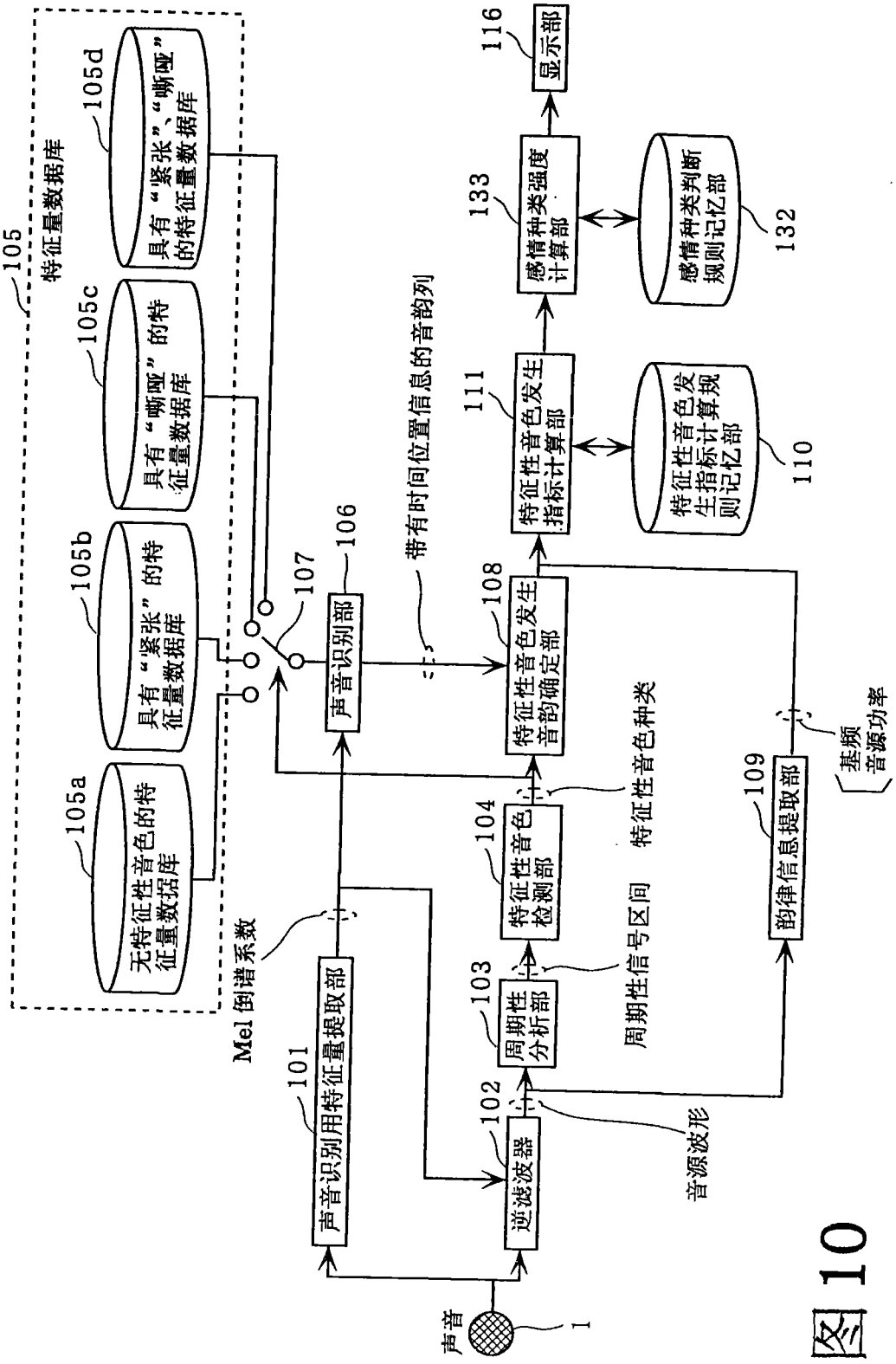

105
特征量数据库
105a
无特征性音色的特征量数据库
105b
具有“紧张”的特征量数据库
105c
具有“嘶哑”的特征量数据库
105d
具有“紧张”、“嘶哑”的特征量数据库
Mel 倒谱系数
101
声音识别用特征量提取部
107
106
声音识别部
带有时间位置信息的音韵列
108
特征性音色发生音韵确定部
111
特征性音色发生指标计算部
特征性音色发生指标计算规则记忆部
110
133
感情种类强度计算部
感情种类判断规则记忆部
132
116
显示部
声音
1
102
逆滤波器
音源波形
103
周期性分析部
周期性信号区间
104
特征性音色检测部
特征性音色种类
109
韵律信息提取部
基频
音源功率

图 10

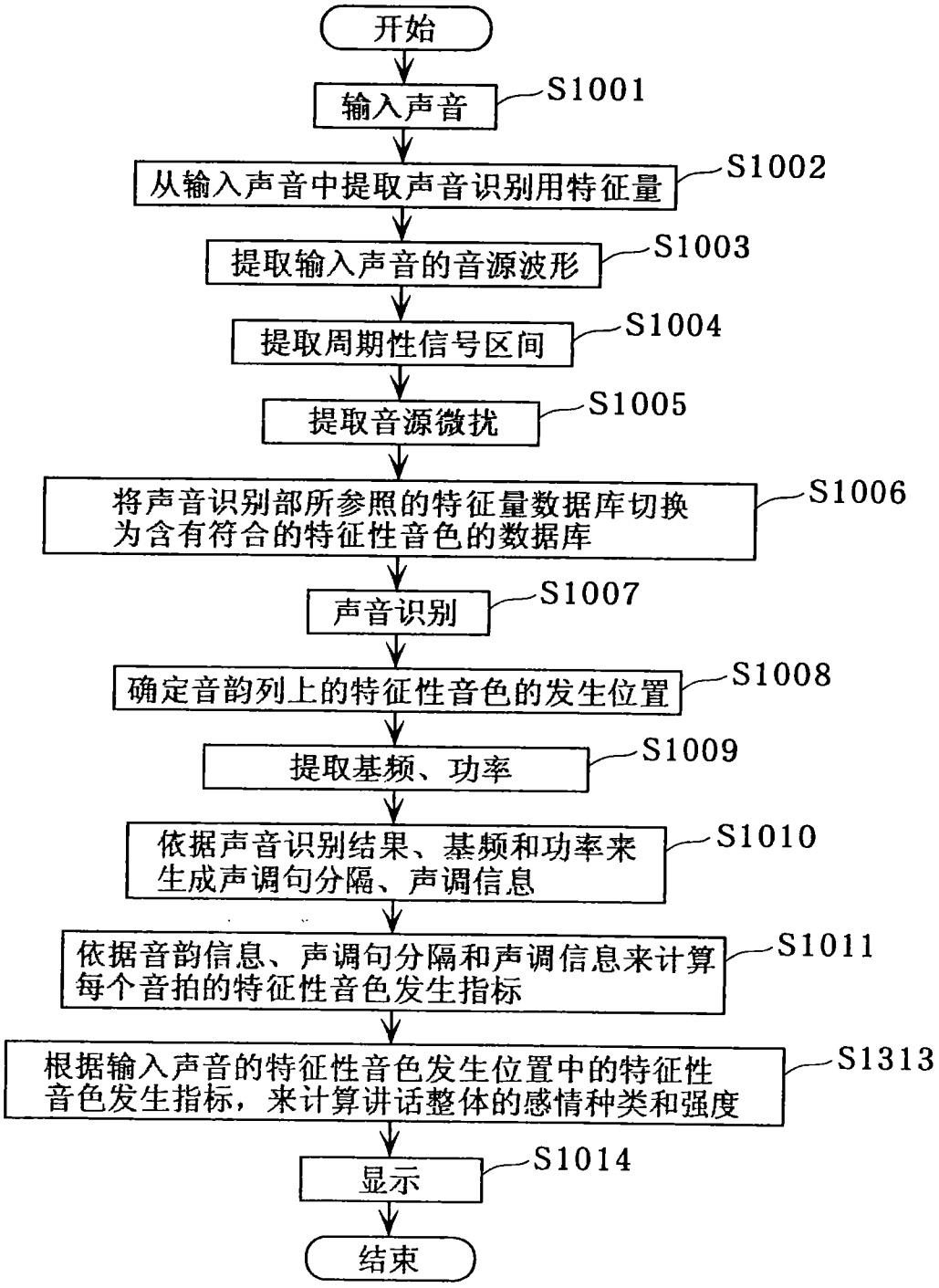

开始
输入声音
S1001
从输入声音中提取声音识别用特征量
S1002
提取输入声音的音源波形
S1003
提取周期性信号区间
S1004
提取音源微扰
S1005
将声音识别部所参照的特征量数据库切换为含有符合的特征性音色的数据库
S1006
声音识别
S1007
确定音韵列上的特征性音色的发生位置
S1008
提取基频、功率
S1009
依据声音识别结果、基频和功率来生成声调句分隔、声调信息
S1010
依据音韵信息、声调句分隔和声调信息来计算每个音拍的特征性音色发生指标
S1011
根据输入声音的特征性音色发生位置中的特征性音色发生指标，来计算讲话整体的感情种类和强度
S1313
显示
S1014
结束

图 11

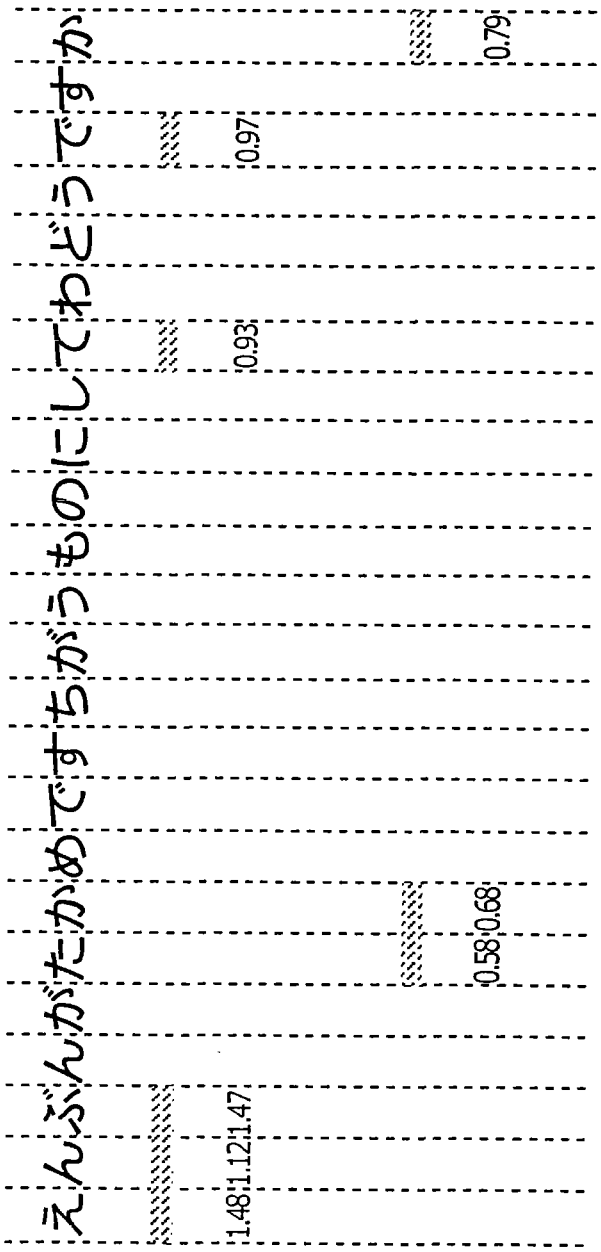
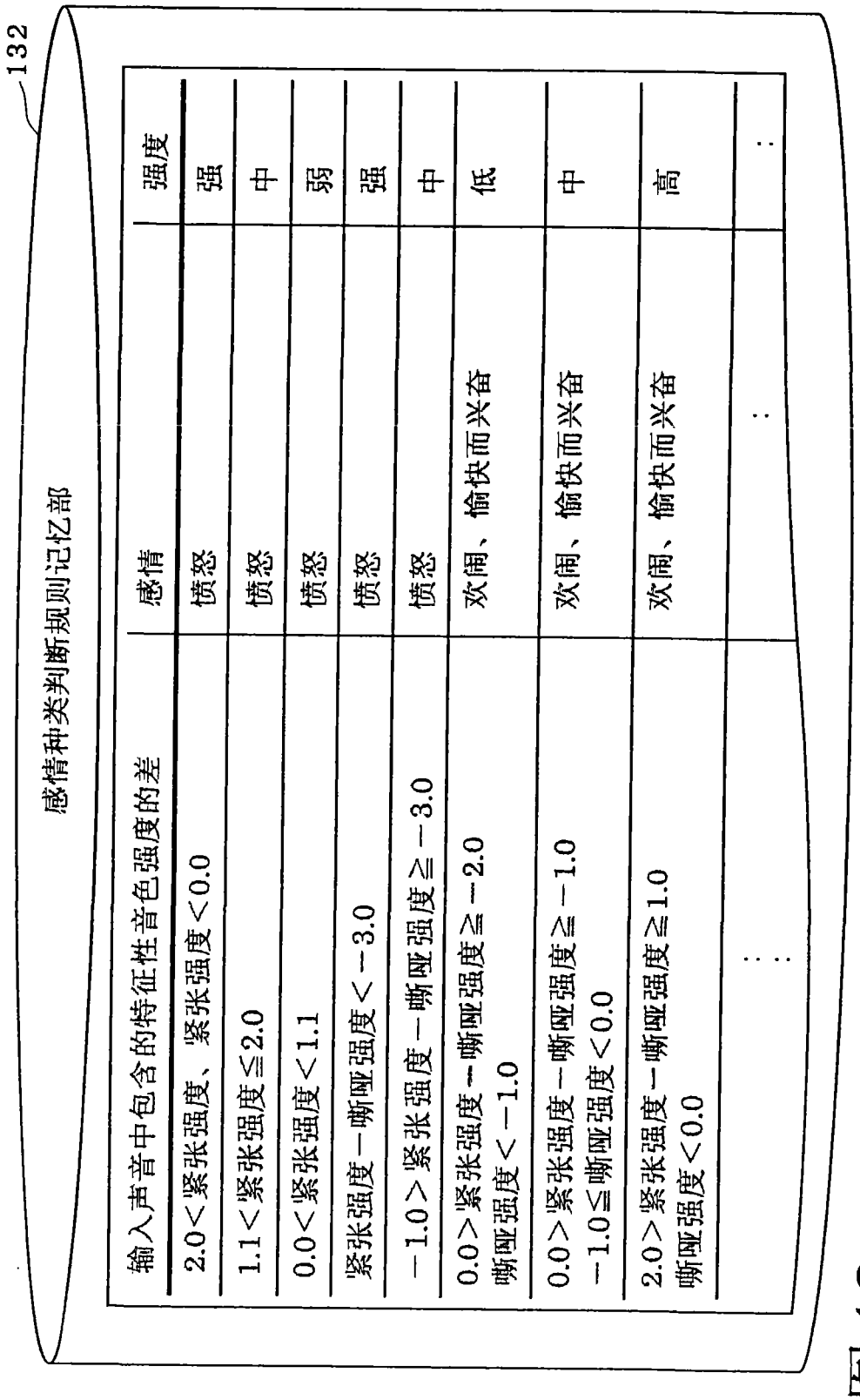

えんぶんがたかめですちがうものにしてわどうですか
1.48
1.12
1.47
0.58
0.68
0.93
0.97
0.79
检测出的紧张
“紧张”发生指标
检测出的嘶哑
“嘶哑”发生指标

图 12



输入声音中包含的特征性音色强度的差	感情	强度
2.0＜紧张强度、紧张强度＜0.0	愤怒	强
1.1＜紧张强度≦2.0	愤怒	中
0.0＜紧张强度＜1.1	愤怒	弱
紧张强度－嘶哑强度＜－3.0	愤怒	强
－1.0＞紧张强度－嘶哑强度≧－3.0	愤怒	中
0.0＞紧张强度－嘶哑强度≧－2.0 嘶哑强度＜－1.0	欢闹、愉快而兴奋	低
0.0＞紧张强度－嘶哑强度≧－1.0 －1.0≦嘶哑强度＜0.0	欢闹、愉快而兴奋	中
2.0＞紧张强度－嘶哑强度≧1.0 嘶哑强度＜0.0	欢闹、愉快而兴奋	高
：：	：：	：：

图 13

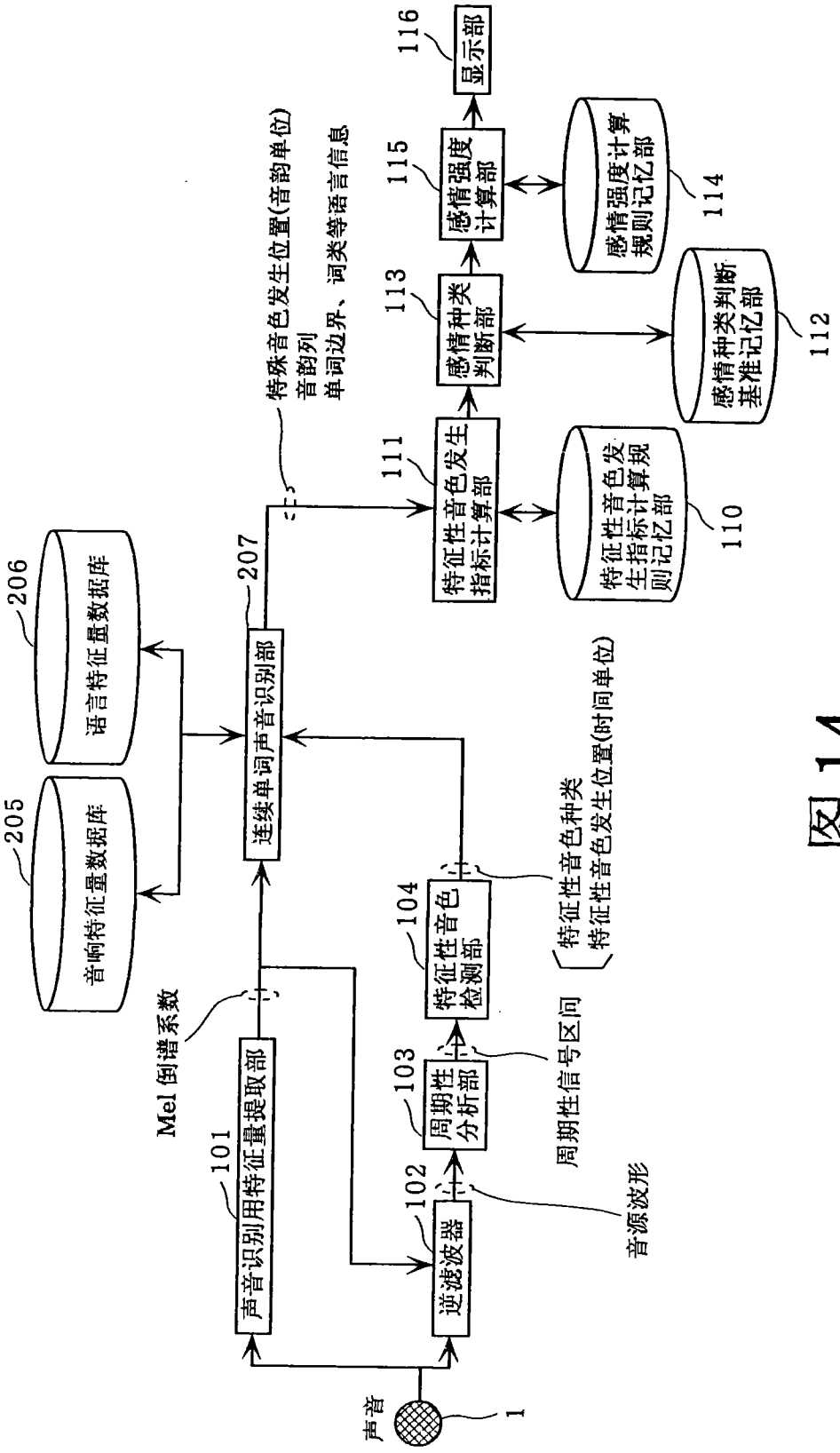

声音
1
逆滤波器
102
声音识别用特征量提取部
101
周期性分析部
103
特征性音色检测部
104
连续单词声音识别部
207
音响特征量数据库
205
语言特征量数据库
206
Mel 倒谱系数
音源波形
周期性信号区间
特征性音色种类
特征性音色发生位置(时间单位)
特殊音色发生位置(音韵单位)
音韵列
单词边界、词类等语言信息
特征性音色发生指标计算部
111
特征性音色发生指标计算规则记忆部
110
感情种类判断部
113
感情种类判断基准记忆部
112
感情强度计算部
115
感情强度计算规则记忆部
114
显示部
116

图 14

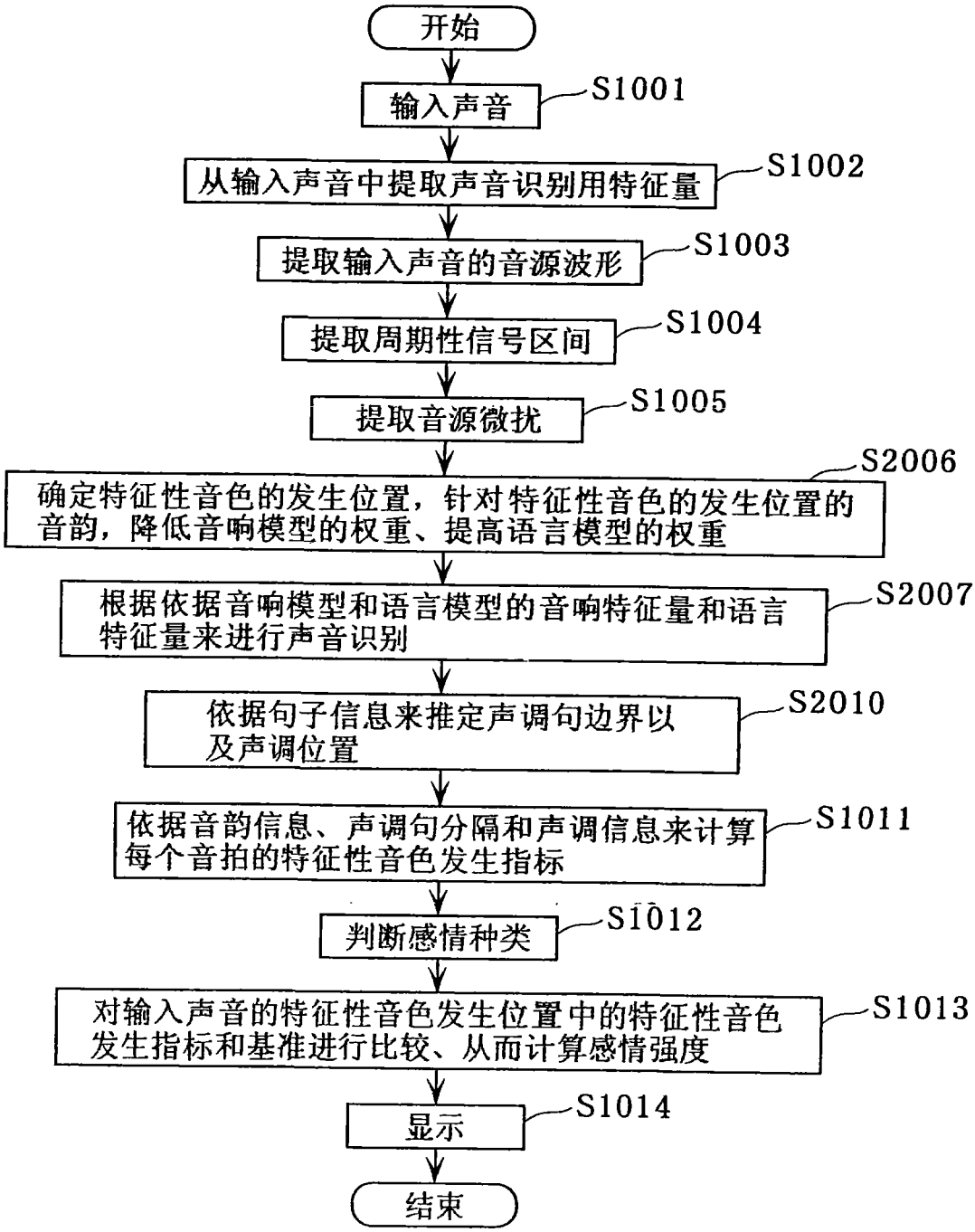

开始
输入声音
S1001
从输入声音中提取声音识别用特征量
S1002
提取输入声音的音源波形
S1003
提取周期性信号区间
S1004
提取音源微扰
S1005
确定特征性音色的发生位置，针对特征性音色的发生位置的音韵，降低音响模型的权重、提高语言模型的权重
S2006
根据依据音响模型和语言模型的音响特征量和语言特征量来进行声音识别
S2007
依据句子信息来推定声调句边界以及声调位置
S2010
依据音韵信息、声调句分隔和声调信息来计算每个音拍的特征性音色发生指标
S1011
判断感情种类
S1012
对输入声音的特征性音色发生位置中的特征性音色发生指标和基准进行比较、从而计算感情强度
S1013
显示
S1014
结束

图 15

なまえをかくえんぴつがほしいんです

“紧张”检测区间

图16A

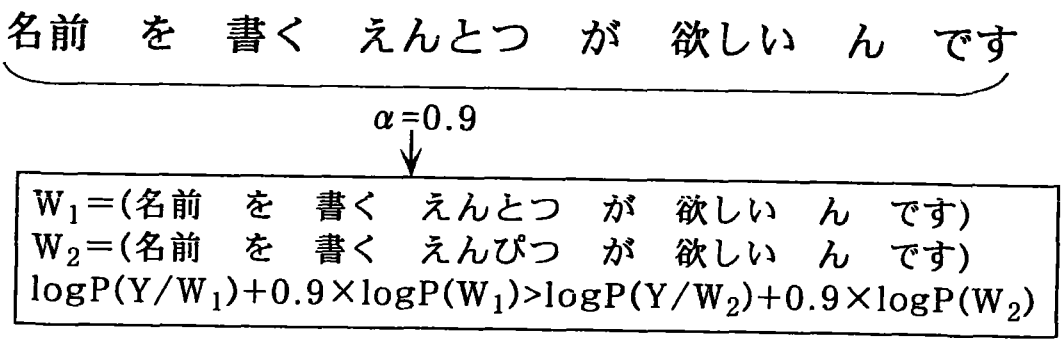


图16B

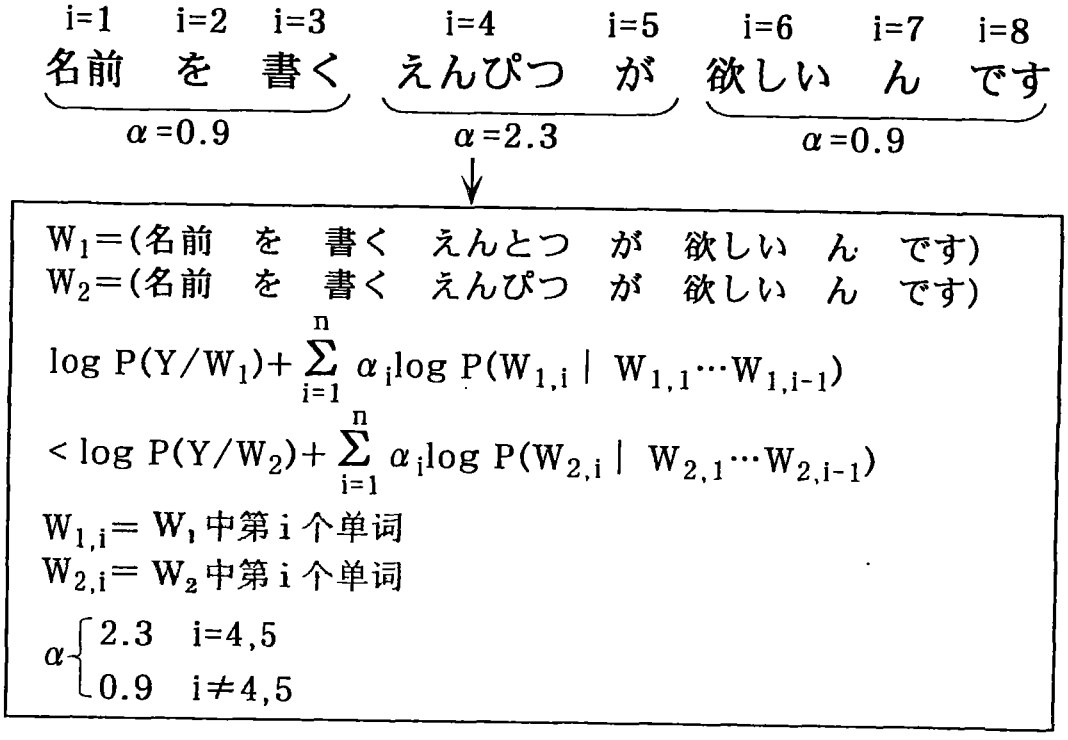


图16C

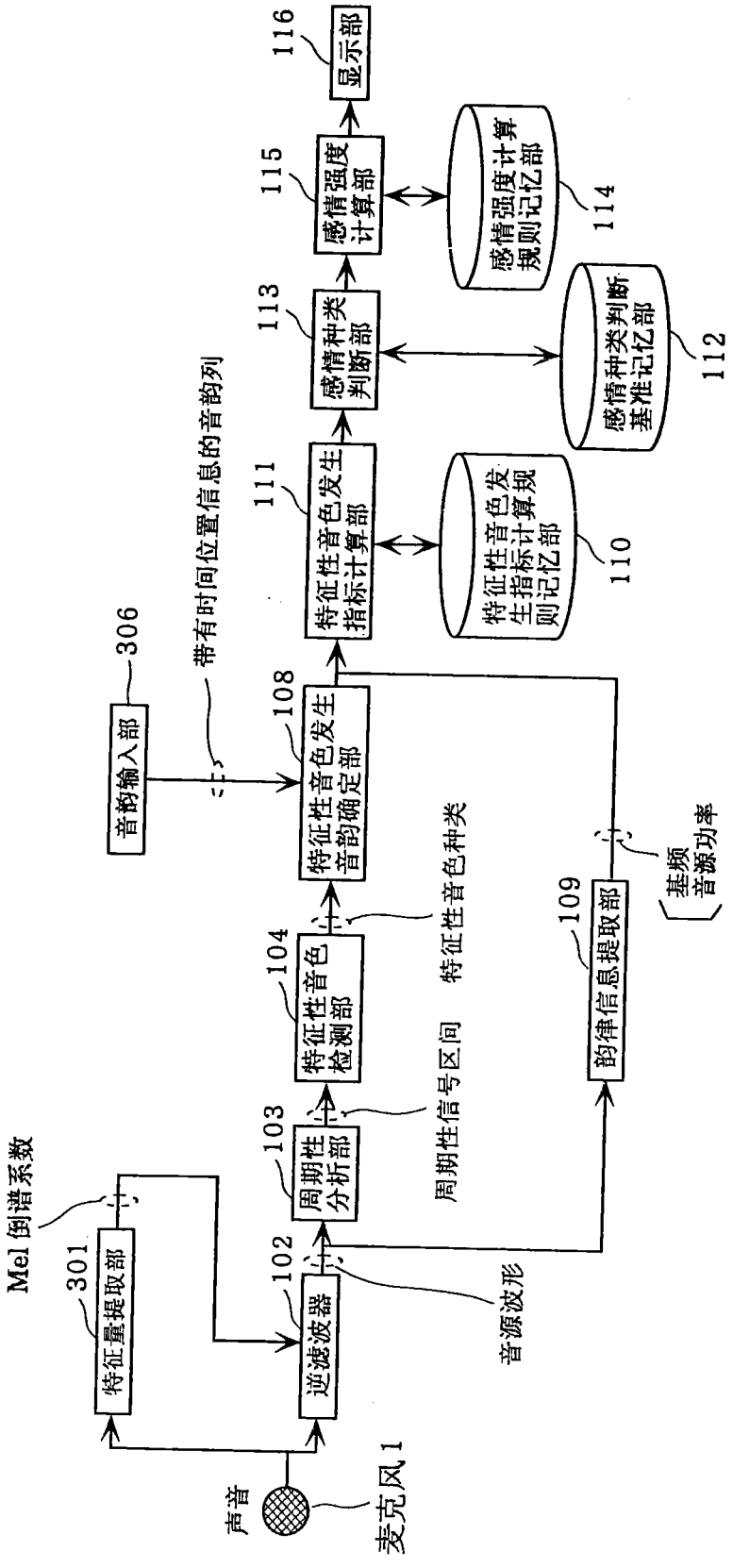

麦克风 1
声音
逆滤波器
102
音源波形
特征量提取部
301
Mel 倒谱系数
周期性分析部
103
周期性信号区间
特征性音色检测部
104
特征性音色种类
特征性音色发生音韵确定部
108
音韵输入部
306
带有时间位置信息的音韵列
韵律信息提取部
109
基频
音源功率
特征性音色发生指标计算部
111
特征性音色发生指标计算规则记忆部
110
感情种类判断部
113
感情种类判断基准记忆部
112
感情强度计算部
115
感情强度计算规则记忆部
114
显示部
116

图 17

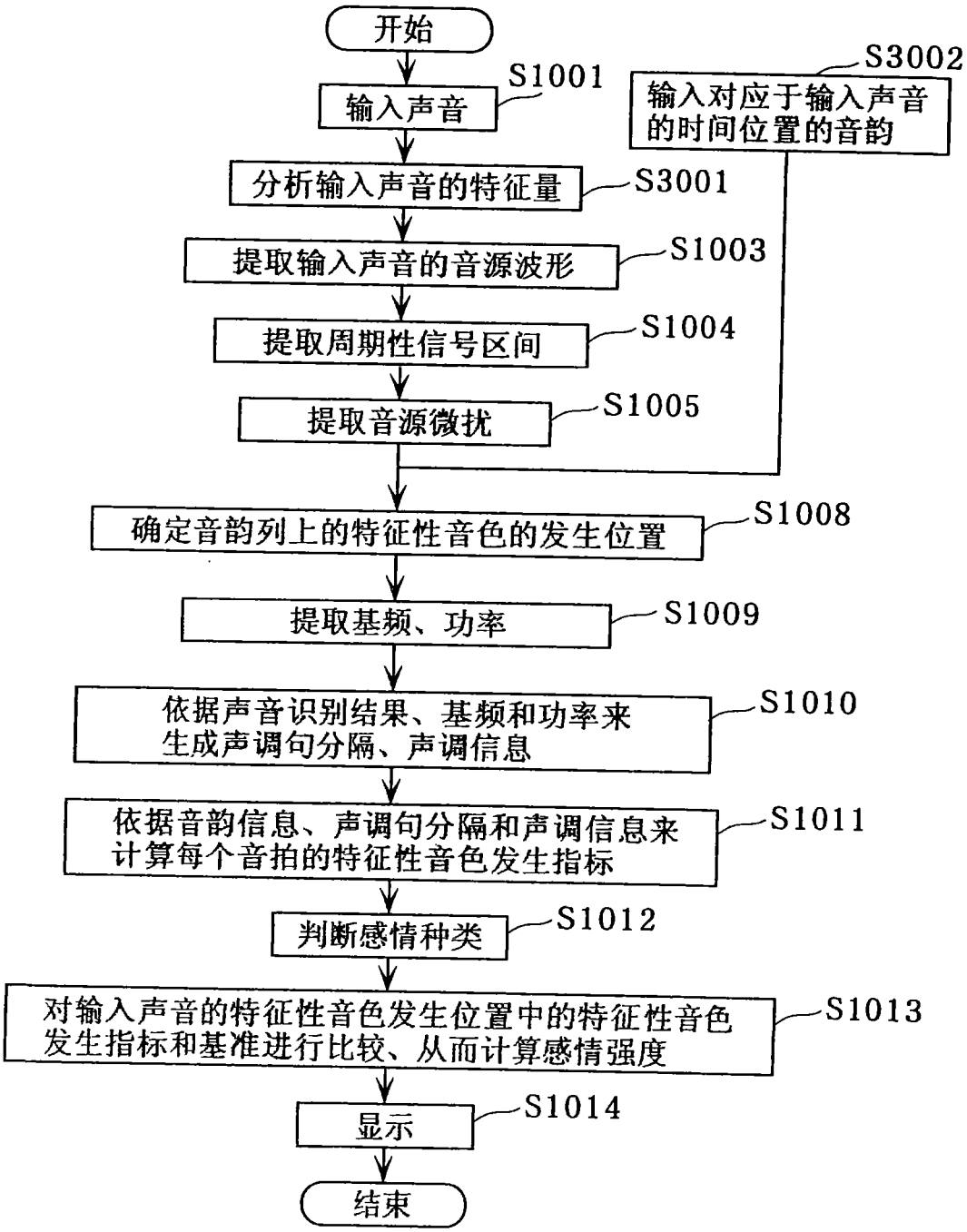

开始
S1001
输入声音
S3002
输入对应于输入声音的时间位置的音韵
分析输入声音的特征量
S3001
提取输入声音的音源波形
S1003
提取周期性信号区间
S1004
提取音源微扰
S1005
确定音韵列上的特征性音色的发生位置
S1008
提取基频、功率
S1009
依据声音识别结果、基频和功率来生成声调句分隔、声调信息
S1010
依据音韵信息、声调句分隔和声调信息来计算每个音拍的特征性音色发生指标
S1011
判断感情种类
S1012
对输入声音的特征性音色发生位置中的特征性音色发生指标和基准进行比较、从而计算感情强度
S1013
显示
S1014
结束

图 18

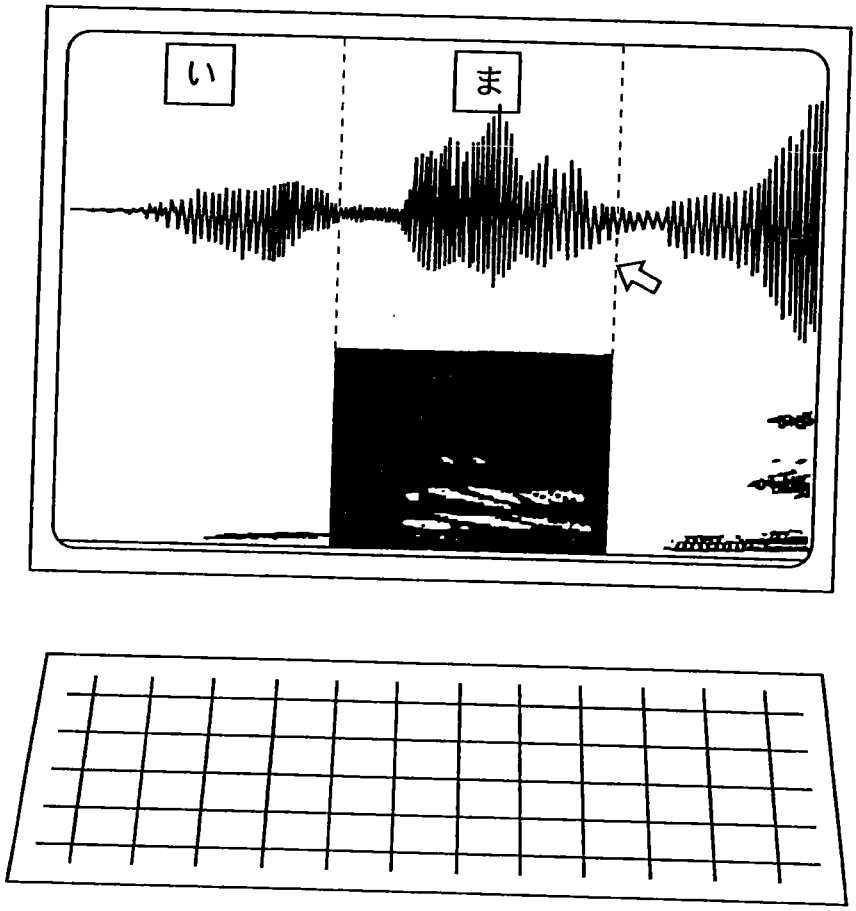


图19

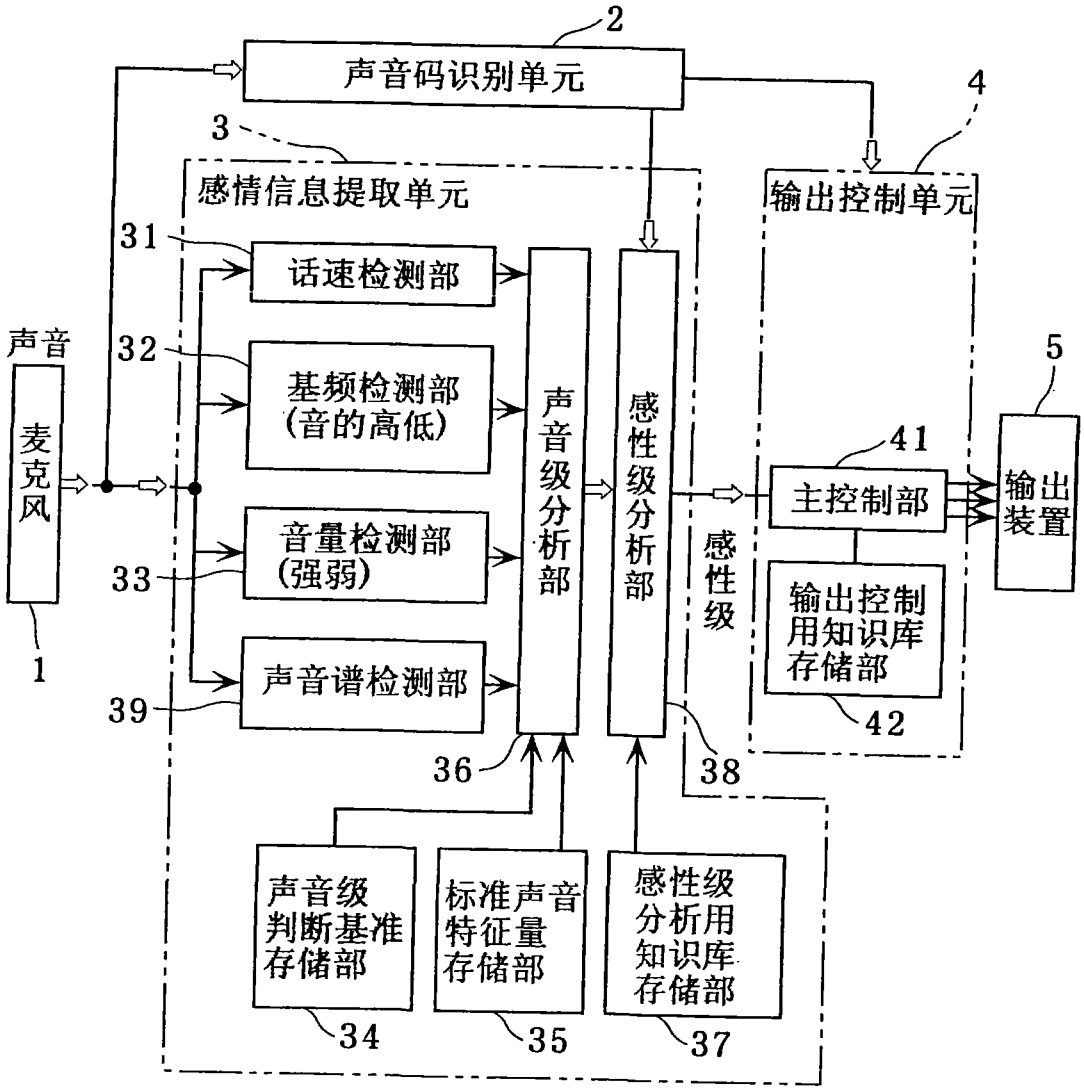

2
声音码识别单元
4
3
感情信息提取单元
输出控制单元
31
话速检测部
声音
32
基频检测部
(音的高低)
声音级分析部
感性级分析部
5
麦克风
41
主控制部
输出装置
感性级
33
音量检测部
(强弱)
输出控制用知识库存储部
1
39
声音谱检测部
42
36
38
声音级判断基准存储部
标准声音特征量存储部
感性级分析用知识库存储部
34
35
37

图 20