[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G01L 13/06 (2006.01)
G10L 13/08 (2006.01)

[12] 发明专利申请公开说明书

[21] 申请号 200580000891.X

[43] 公开日 2006 年 10 月 4 日

[11] 公开号 CN 1842702A

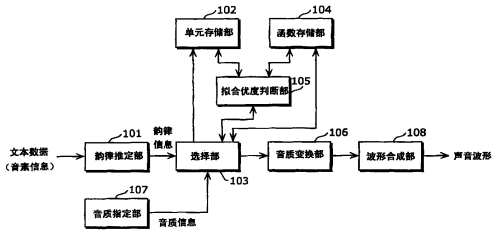
[22] 申请日 2005.9.20
[21] 申请号 200580000891.X
[30] 优先权
[32] 2004.10.13 [33] JP [31] 299365/2004
[32] 2005. 7. 7 [33] JP [31] 198926/2005
[86] 国际申请 PCT/JP2005/017285 2005.9.20
[87] 国际公布 WO2006/040908 日 2006.4.20
[85] 进入国家阶段日期 2006.3.15
[71] 申请人 松下电器产业株式会社
地址 日本大阪府
[72] 发明人 广濑良文 斋藤夏树 釜井孝浩
[74] 专利代理机构 永新专利商标代理有限公司
代理人 黄剑锋
权利要求书 5 页 说明书 41 页 附图 31 页

[54] 发明名称

声音合成装置和声音合成方法

[57] 摘要

提供可适当变换音质的声音合成装置。该声音合成装置具有：单元存储部(102)，存储有多个声音单元；函数存储部(104)，存储有多个变换函数；拟和优度判断部(105)，比较单元存储部(102)中存储的声音单元、和制作函数存储部(104)中存储的变换函数时使用的声音单元的音响特征，来导出类似度；选择部(103)和音质变换部(106)，根据由拟和优度判断部(105)导出的类似度，对单元存储部(104)中存储的每个声音单元，应用函数存储部(104)中存储的某个变换函数，从而变换该声音单元的音质。



1、一种声音合成装置，利用声音单元合成声音，以变换音质，其特征在于，具有：

单元存储机构，存储有多个声音单元；

函数存储机构，存储有用于变换声音单元的音质的多个变换函数；

类似度导出机构，比较表示上述单元存储机构中存储的声音单元的音响特征、和制作上述函数存储机构中所存储的变换函数时使用的声音单元的音响特征，来导出类似度；

变换机构，根据由上述类似度导出机构导出的类似度，对上述单元存储机构中存储的每个声音单元，应用上述函数存储机构中存储的某个变换函数，从而变换该声音单元的音质。

2、如权利要求 1 所述的声音合成装置，其特征在于，

上述单元存储机构中存储的声音单元的声音特征和制作上述变换函数时使用的声音单元的声音特征越类似，上述类似度导出机构导出越高的类似度；

上述变换机构对上述单元存储机构中存储的声音单元，应用使用上述类似度最高的声音单元来制作的变换函数。

3、如权利要求 2 所述的声音合成装置，其特征在于，

上述类似度导出机构根据由上述单元存储机构中存储的声音单元和该声音单元的前后声音单元构成的系列音响特征、和由制作上述变换函数时使用的声音单元及该声音单元的前后声音单元构成的系列音响特征之间的类似度，来导出动态的上述类似度。

4、如权利要求 2 所述的声音合成装置，其特征在于，

上述类似度导出机构根据上述单元存储机构中存储的声音单元的音响特征和制作上述变换函数时使用的声音单元的音响特征之间的类似度，来导出静态的上述类似度。

5、如权利要求 1 所述的声音合成装置，其特征在于，

上述变换机构对上述单元存储机构中存储的声音单元，应用使用上述类似度大于等于规定阈值的声音单元来制作的变换函数。

6、如权利要求 1 所述的声音合成装置，其特征在于，

上述声音合成装置还具有生成机构，该生成机构生成表示对应于用户操作的音素和韵律的韵律信息；

上述变换机构具有：

选择机构，根据上述类似度，从上述单元存储机构及函数存储机构相辅地选择对应于上述韵律信息所表示的音素及韵律的声音单元、和对应于上述韵律信息所表示的音素及韵律的变换函数；以及

应用机构，对由上述选择机构选择的声音单元应用由上述选择机构选择的变换函数。

7、如权利要求 6 所述的声音合成装置，其特征在于，

上述声音合成装置还具有接受由用户指定的音质的音质指定机构；

上述选择机构选择用于变换成由上述音质指定机构接受的音质的变换函数。

8、如权利要求 6 所述的声音合成装置，其特征在于，

上述生成机构根据用户操作取得文本数据，并根据包含在上述文本数据中的音素来推定韵律，生成上述韵律信息。

9、如权利要求 1 所述的声音合成装置，其特征在于，

上述声音合成装置还具有生成表示对应于用户操作的音素及韵律的韵律信息的生成机构；

上述变换机构具有：

函数选择机构，从上述函数存储机构选择对应于上述韵律信息所表示的音素及韵律的变换函数；

单元选择机构，对由上述函数选择机构选择的变换函数，根据上述类似度、从上述单元存储机构选择对应于上述韵律信息所表示的音素及韵律的声音单元；以及

应用机构，对由上述单元选择机构选择的声音单元，应用由上述函数选择机构选择的变换函数。

10、如权利要求 1 所述的声音合成装置，其特征在于，

上述声音合成装置还具有生成表示对应于用户操作的音素及韵律的韵律信息；

上述变换机构具有：

单元选择机构，从上述单元存储机构选择对应于上述韵律信息所表示的音素及韵律的声音单元；

函数选择机构，对由上述单元选择机构选择的声音单元，根据上述类似度、从上述函数存储机构选择对应于上述韵律信息所表示的音素及韵律的变换函数；以及

应用机构，对由上述单元选择机构选择的声音单元，应用由上述函数选择机构选择的变换函数。

11、如权利要求 1 所述的声音合成装置，其特征在于，

上述单元存储机构存储着构成第 1 音质的声音的多个声音单元；

上述函数存储机构对第 1 音质的声音的每个声音单元，将该声音单元、表示该声音单元的音响特征的基准代表值、和对上述基准代表值的变换函数分别关联起来进行存储；

上述声音合成装置还具有代表值确定机构，该代表值确定机构对上述单元存储机构中存储的第 1 音质的声音的每个声音单元，确定表示该声音单元的音响特征的代表值；

上述类似度导出机构比较上述单元存储机构中存储的声音单元所表示的上述代表值、和制作上述函数存储机构中存储的变换函数时使用的声音单元的上述基准代表值，来导出类似度；

上述变换机构具有：

选择机构，对上述单元存储机构中存储的每个声音单元，从同与该声音单元相同的声音单元相关联地存储在上述函数存储装置中的变换函数中，选择与和该声音单元的代表值的类似度最高的基准代表值相关联

的变换函数；

函数应用机构，对上述单元存储机构中存储的声音单元，通过将由上述选择机构选择的变换函数应用于上述声音单元，来将上述第 1 音质的声音变换为第 2 音质的声音。

12、如权利要求 11 所述的声音合成装置，其特征在于，

上述声音合成装置还具有声音合成机构，该声音合成机构取得文本数据，并生成表示与上述文本数据相同内容的上述多个声音单元，存储到上述单元存储机构中。

13、如权利要求 12 所述的声音合成装置，其特征在于，

上述声音合成机构具有：

单元代表值存储机构，将构成上述第 1 音质的声音的各声音单元和表示上述各声音单元的音响特征的代表值相关联起来进行存储；

分析机构，取得并分析上述文本数据；

选择存储机构，根据上述分析机构的分析结果，从上述单元代表值存储机构选择对应于上述文本数据的声音单元，并将所选择的声音单元和该声音单元的代表值关联起来存储到上述单元存储机构中；

上述代表值确定机构对上述单元存储机构中存储的每个声音单元，确定与该声音单元关联起来存储的代表值。

14、如权利要求 13 所述的声音合成装置，其特征在于，

上述声音合成装置还具有：

基准代表值存储机构，对上述第 1 音质的声音的每个声音单元，存储着该声音单元和表示该声音单元的音响特征的基准代表值；

目标代表值存储机构，对上述第 2 音质的声音的每个声音单元，存储着该声音单元和表示该声音单元的音响特征的目标代表值；

变换函数生成机构，根据与上述基准代表值存储机构和目标代表值存储机构中存储的相同的声音单元对应的基准代表值和目标代表值，生成对上述基准代表值的上述变换函数。

15、如权利要求 14 所述的声音合成装置，其特征在于，

上述声音单元是音素，表示上述音响特征的代表值和基准代表值分别是音素的时间中心处的共振峰频率值。

16、如权利要求 14 所述的声音合成装置，其特征在于，

上述声音单元是音素，表示上述音响特征的代表值和基准代表值分别是音素的共振峰频率的平均值。

17、一种声音合成方法，利用声音单元合成声音，以变换音质，其特征在于，

单元存储机构存储有多个声音单元，函数存储机构存储有用于变换声音单元的音质的多个变换函数，

上述声音合成方法包括：

类似度导出步骤，比较上述单元存储机构中存储的声音单元所表示的音响特征、和制作上述函数存储机构中存储的变换函数时使用的声音单元的音响特征，来导出类似度；

变换步骤，根据由上述类似度导出机构导出的类似度，对上述单元存储机构中存储的每个声音单元，应用上述函数存储机构中存储的某个变换函数，从而变换该声音单元的音质。

18、一种程序，利用声音单元合成声音，以变换音质，其特征在于，

单元存储机构存储有多个声音单元，函数存储机构存储有用于变换声音单元的音质的多个变换函数，

上述程序使计算机执行以下步骤：

类似度导出步骤，比较上述单元存储机构中存储的声音单元所表示的音响特征、和制作上述函数存储机构中存储的变换函数时使用的声音单元的音响特征，来导出类似度；

变换步骤，根据由上述类似度导出步骤导出的类似度，对上述单元存储机构中存储的每个声音单元，应用上述函数存储机构中存储的某个变换函数，从而变换该声音单元的音质。

声音合成装置和声音合成方法

技术领域

本发明涉及利用声音单元合成声音的声音合成装置和声音合成方法，尤其涉及变换音质的声音合成装置和声音合成方法。

背景技术

在现有技术中，已经提出有变换音质的声音合成装置，例如参照专利文献 1~3。

专利文献 1：日本特开平 7-319495 号公报（第 0014 段落至第 0019 段落）；

专利文献 2：日本特开 2003-66982 号公报（第 0035 段落至第 0053 段落）；

专利文献 3：日本特开 2002-215198 号公报。

上述专利文献 1 的声音合成装置通过保持不同音质的多个声音单元组、并切换使用声音单元组，来进行音质的变换。

图 1 是表示上述专利文献 1 的声音合成装置结构的结构图。

该声音合成装置包括合成单位数据信息表 901、个人代码簿保存部 902、 似然计算部 903、多个个人合成单位数据库 904、音质变换部 905。

合成单位数据信息表 901 保持与作为声音合成对象的合成单位有关的数据（合成单位数据）。在这些合成单位数据中，分配有用于识别各合成单位数据的合成单位数据 ID。个人代码簿保存部 902 存储所有讲话者的标识符（个人标识 ID）和表示其音质特征的信息。似然计算部 903 根据基准参数信息、合成单位名称、音韵环境信息、目标音质信息，并参考合成单位数据信息表 901 和个人代码簿保存部 902，来选择合成单位数

据 ID 和个人标识 ID。

多个个人合成单位数据库 904 保持音质互不相同的声音单元组。并且，各个人合成单位数据库 904 与个人标识 ID 相对应。

音质变换部 905 取得由似然计算部 903 选择的合成单位数据 ID 和个人标识 ID。并且，音质变换部 905 从该个人表示 ID 所表示的个人合成单位数据库 904 取得与表示该合成单位数据 ID 所表示的合成单位数据对应的声音单元，来生成声音波形。

另一方面，上述专利文献 2 的声音合成装置通过使用用于进行音质变换的变换函数，来变换通常的合成音的音质。

图 2 是表示上述专利文献 2 的声音合成装置的结构图。

该声音合成装置包括文本输入部 911、单元存储部 912、单元选择部 913、音质变换部 914、波形合成部 915、音质变换参数输入部 916。

文本输入部 911 取得表示要合成的语言内容的文本信息或音素信息、和表示重音或讲话整体的抑扬的韵律信息。单元存储部 912 存储一组声音单元（合成声音单位）。单元选择部 913 根据由文本输入部 911 取得的音素信息或韵律信息，从单元存储部 912 选择多个最佳声音单元，并输出该选择的多个声音单元。音质变换参数输入部 916 取得表示有关音质的参数的音质参数。

音质变换部 914 根据由音质变换参数输入部 916 取得的音质参数，对由单元选择部 913 选择的声音单元进行音质变换。从而对该声音单元进行线形或非线性的频率变换。波形合成部 915 根据由音质变换部 914 进行了音质变换的声音单元，生成声音波形。

图 3 是用于说明在上述专利文献 2 的声音变换部 914 中的声音单元的音质变换中使用的变换函数的说明图。在此，图 3 的横轴（Fi）表示输入到音质变换部 914 的声音单元的输入频率，图 3 的纵轴（Fo）表示由音质变换部 914 输出的声音单元的输出频率。

在作为音质参数使用变换函数 f101 的情况下，音质变换部 914 不对由单元选择部 913 选择的声音单元进行音质变换就输出。此外，在作为

音质参数使用变换函数 f102 的情况下，音质变换部 914 对由单元选择部 913 选择的声音单元的输入频率进行线性变换之后输出，并在作为音质参数使用变换函数 f103 的情况下，对由单元选择部 913 选择的声音单元的输入频率进行非线性变换之后输出。

此外，专利文献 3 的声音合成装置（音质变换装置）根据音质变换对象的音素的音响特征，来判断属于该音素群。并且，该声音合成装置利用对属于该音素的群设定的变换函数来变换该音素的音质。

但是，在上述专利文献 1~专利文献 3 的声音合成装置中，存在不能变换为适当的音质的问题。

即，上述专利文献 1 的声音合成装置由于切换个人合成单位数据库 904 来变换合成音的音质，所以不能进行连续的音质变换，或不能生成在各个人合成单位数据库 904 中没有的音质的声音波形。

此外，上述专利文献 2 的声音合成装置由于对表示文本信息的输入文整体进行音质变换，因而不能对各音韵进行最佳变换。并且，由于专利文献 2 的声音合成装置依次且独立地进行声音单元的选择和音质变换，如图 3 所示，通过变换函数 f102，有时共振峰频率（输出频率 F0）超过奈奎斯特频率（Nyquist frequency）fn。这种情况下，专利文献 2 的声音合成装置盲目地对共振峰频率进行校正而使其小于等于奈奎斯特频率 fn。其结果，不能变换为适当的音质。

此外，由于上述专利文献 3 的声音合成装置对属于组的所有音素使用相同的变换函数，因此有时在变换后的声音中产生变形。即，对各音素的组划分是根据各音素的音响特征是否满足对各组设定的阈值来进行。在这种情况下，若对充分满足某个组的阈值的音素应用该组的变换函数，则该音素的音质被适当变换。但是，如果对音响特征存在于某个组的阈值附近的音素应用该组的变换函数，则该音素变换后的音质中产生变形。

发明内容

在此，本发明是鉴于上述问题而做出的，其目的在于可适当变换音质的声音合成装置和声音合成方法。

为了达到上述目的，本发明的声音合成装置，利用声音单元合成声音，以变换音质，其特征在于，具有：单元存储机构，存储有多个声音单元；函数存储机构，存储有用于变换声音单元的音质的多个变换函数；类似度导出机构，比较表示上述单元存储机构中所存储的声音单元的音响特征、和制作上述函数存储机构中所存储的变换函数时使用的声音单元的音响特征，来导出类似度；变换机构，根据由上述类似度导出机构导出的类似度，对上述单元存储机构中存储的每个声音单元应用上述函数存储机构中所存储的几个变换函数，从而变换该声音单元的音质。例如，上述类似度导出机构导出上述单元存储机构中存储的声音单元的声音特征与制作上述变换函数时使用的声音单元的声音特征类似的程度高的类似度；上述变换机构对上述单元存储机构中存储的声音单元应用使用上述类似度最高的声音单元来制作的变换函数。此外，上述声音特征是倒频谱距离（Cepstrum Distance）、共振峰频率、基本频率、持续时间长度和功率中的至少一个。

从而，由于用变换函数变换音质，所以能够连续变换音质，并且，对每个声音单元根据类似度来应用变换函数，因此，能够对各声音单元进行最佳的变换。并且，不像现有例那样不需要在变换后进行用于将共振峰频率抑制在规定范围内的无理的校正，即可适当变换音质。

在此，上述声音合成装置还具有生成表示对应于用户操作的音素和韵律的韵律信息的生成机构；上述变换机构具有：选择机构，根据上述类似度，从上述单元存储机构及函数存储机构相辅地选择对应于上述韵律信息表示的音素及韵律的声音单元、和对应于上述韵律信息表示的音素及韵律的变换函数；应用机构，对由上述选择机构选择的声音单元应用由上述选择机构选择的变换函数。

从而，根据类似度来选择由韵律信息表示的音素及对应于韵律的声音单元和变换函数，并将变换函数应用于该声音单元，因此，可通过改

变韵律信息的内容，能够对所希望的音素及韵律次变换音质。此外，由于根据类似度来相辅地选择声音单元及变换函数，所以能够更适当地变换音质。

此外，上述声音合成装置还具有生成表示对应于用户操作的音素及韵律的韵律信息的生成机构；上述变换机构具有：函数选择机构，从上述函数存储机构选择对应于表示上述韵律信息的音素及韵律的变换函数；单元选择机构，对由上述函数选择机构选择的变换函数，根据上述类似度从上述单元存储机构选择对应于表示上述韵律信息的音素及韵律的声音单元；应用机构，对由上述单元选择机构选择的声音单元，应用由上述函数选择机构选择的变换函数。

从而，首先选择对应于韵律信息的变换函数，由于对于该变换函数，根据类似度来选择声音单元，所以，例如即使函数存储单元中存储的变换函数的个数较少，只要单元存储机构中存储的声音单元的个数多，就能够适当变换音质。

上述声音合成装置还具有生成表示对应于用户操作的音素及韵律的韵律信息；上述变换机构具有：单元选择机构，从上述单元存储机构选择对应于上述韵律信息的音素及韵律的声音单元；函数选择机构，对由上述单元选择机构选择的声音单元，根据上述类似度从上述函数存储机构选择对应于表示上述韵律信息的音素及韵律的变换函数；应用机构，对由上述单元选择机构选择的声音单元应用由上述函数选择机构选择的变换函数。

从而，首先选择对应于韵律信息的变换函数，由于对于该声音单元，根据类似度来选择变换函数，所以，例如即使函数存储单元中存储的声音单元的个数较少，只要单元存储机构中存储的变换函数的个数多，就能够适当变换音质。

在此，上述声音合成装置还具有接受由用户指定的音质的音质指定机构；上述选择机构选择用于变换为由上述音质指定机构接受的音质的变换函数。

从而，由于用于变换为由用户指定的音质的变换函数被选择，因此能够适当地变换为所希望的音质。

在此，上述类似度导出机构根据由上述单元存储机构中存储的声音单元和该声音单元的前后声音单元构成的一系列音响特征、和由制作上述变换函数时使用的声音单元及该声音单元的前后声音单元构成的一系列音响特征之间的类似度，来导出动态的上述类似度。

从而，由于使用与由单元存储机构的系列整体表示的音响特征类似的系列来制作的变换函数，应用于该单元存储机构的系列中包含的声音单元，因此能够确保该系列整体的音质的调和。

再有，上述单元存储机构存储构成第 1 音质的声音的多个声音单元；上述函数存储机构对地 1 音质的声音的声音单元，将该声音单元、表示该声音单元的音响特征的基准代表值、和对上述基准代表值的变换函数分别关联起来进行存储；上述声音合成装置还具有代表值确定机构，该代表值确定机构对上述单元存储机构中存储的第 1 音质的声音的声音单元，确定表示该声音单元的音响特征的代表值；上述类似度导出机构比较表示上述单元存储机构中存储的声音单元的上述代表值和制作上述函数存储机构中存储的变换函数时使用的声音单元的上述基准代表值，来导出类似度。上述变换机构具有：选择机构，对上述单元存储机构中存储的每个声音单元，从与该声音单元相同的声音单元相关联地存储在上述函数存储装置中的变换函数中，选择与和该声音单元的代表值的类似度最高的基准代表值相关联的变换函数；函数应用机构，对上述单元存储机构中存储的声音单元，通过将由上述选择机构选择的变换函数应用于上述声音单元，来将上述第 1 音质的声音变换为第 2 音质的声音。

从而，在对第 1 音质的声音的音素选择变换函数时，不像现有例那样与该音素的音响特征无关地对该音素选择预先设定的变换函数，而选择与该音素的音响特征所表示的代表值最近的基准代表值关联的变换函数。因此，即使是同一音素其频谱（音响特征）根据上下文或感情而变动，但是在本发明中，能够进行使用了对该频谱所具有的音素总是最佳

的变换函数的音质变换，能够适当变换音质。即，为了保证变换后的频谱的妥当性，能够得到高质量的音质变换声音。

此外，本发明中，用代表值和基准代表值简单地表示音响特征，所以在从函数存储机构选择变换函数时，能够不进行复杂的运算处理而简单且迅速和适当地选择变换函数。例如，在用频谱表示音响特征时，必须通过复杂的处理比较地 1 音质的音素的频谱和函数存储机构的音素的频谱，但是本发明中能够减轻这样的处理负担。此外，由于在函数存储机构中作为音响特征而存储有基准代表值，所以与作为音响特征而存储频谱的情况相比，能够减小函数存储机构的存储容量。

在此，上述声音合成装置还具有声音合成机构，该声音合成机构取得文本数据，并生成表示与上述文本数据相同的内容的上述多个声音单元之后，存储到上述单元存储机构中。

此时，上述声音合成机构具有：单元代表值存储机构，将构成上述第 1 音质的声音的各声音单元和表示上述各声音单元的音响特征的代表值相关联起来进行存储；分析机构，取得并分析上述文本数据；选择存储机构，根据上述分析机构的分析结果，从上述单元代表值存储机构选择对应于上述文本数据的声音单元，并将所选择的声音单元和该声音单元的代表值向关联起来存储到上述单元存储机构中；上述代表值确定机构对上述单元存储机构中存储的每个声音单元，确定与该声音单元关联起来存储的代表值。

从而，通过将文本数据经第 1 音质的声音适当地变换为第 2 音质的声音。

此外，上述声音合成装置还具有：基准代表值存储机构，对上述第 1 音质的声音的每个声音单元，存储该声音单元和表示该声音单元的音响特征的基准代表值；目标代表值存储机构，对上述第 2 音质的声音的每个声音单元，存储该声音单元和表示该声音单元的音响特征的目标代表值；变换函数生成机构，根据与上述基准代表值存储机构和目标代表值存储机构中存储的相同的声音单元对应的基准代表值和目标代表值，声

称对上述基准代表值的上述变换函数。

从而，根据表示第 1 音质的音响特征的基准代表值和表示第 2 音质的音响特征的目标代表值来生成变换函数，因此能够防止无理的音质变换的音质的破绽，能够将第 1 音质可靠地变换为第 2 音质。

在此，表示上述音响特征的代表值和基准代表值分别是音素的时间中心的共振峰频率的值。

特别是，由于在元音的时间中心，共振峰频率稳定，所以能够将第 1 音质适当地变换为第 2 音质。

此外，表示上述音响特征的代表值和基准代表值分别是音素的共振峰频率的平均值。

特别是，由于在无声辅音中共振峰频率的平均值适当地表示音响特征，所以能够将第 1 音质适当地变换为第 2 音质。

此外，不仅能够作为上述的声音合成装置来实现，还可以作为合成声音的方法、或使计算机基于该方法来合成声音的程序、存储有该程序的存储介质来实现。

本发明的声音合成装置具有可适当变换音质的作用效果。

附图说明

图 1 是表示专利文献 1 的声音合成装置的结构的结构图。

图 2 是表示专利文献 2 的声音合成装置的结构的结构图。

图 3 是用于说明在专利文献 2 的音质变换部中的声音单元的音质变换中使用的变换函数的说明图。

图 4 是表示本发明的第 1 实施方式中的声音合成装置的结构的结构图。

图 5 是表示同上的选择部的结构的结构图。

图 6 是用于说明同上的单元点阵确定部和函数点阵确定部的动作的说明图。

图 7 是用于说明同上的动态拟合优度的说明图。

图 8 是表示同上的选择部的动作的流程图。

图 9 是表示同上的声音合成装置的动作的流程图。

图 10 是表示元音“i”的声音频谱的图。

图 11 是表示元音“i”的其他声音频谱的图。

图 12A 是表示对元音“i”的频谱应用变换函数的例的图。

图 12B 是表示对元音“i”的其他频谱应用变换函数的例的图。

图 13 是用于说明第 1 实施方式中的声音合成装置适当地选择变换函数的情况的说明图。

图 14 是用于说明有关同上的变形例的单元点阵确定部和函数点阵确定部的动作的说明图。

图 15 是表示本发明的第 2 实施方式中的声音合成装置的结构的结构图。

图 16 是表示同上的函数选择部的结构的结构图。

图 17 是表示同上的单元选择部的结构的结构图。

图 18 是表示同上的声音合成装置的动作的流程图。

图 19 是表示本发明的第 3 实施方式中的声音合成装置的结构的结构图。

图 20 是表示同上的单元选择部的结构的结构图。

图 21 是表示同上的函数选择部的结构的结构图。

图 22 是表示同上的声音合成装置的动作的流程图。

图 23 是表示本发明的第 4 实施方式的音质变换装置(声音合成装置)的结构的结构图。

图 24A 是表示同上的音质 A 的基点信息的一例的示意图。

图 24B 是表示同上的音质 B 的基点信息的一例的示意图。

图 25A 是用于说明同上的 A 基点数据库中存储的信息的说明图。

图 25B 是用于说明同上的 B 基点数据库中存储的信息的说明图。

图 26 是表示同上的函数提取部的处理例的示意图。

图 27 是表示同上的函数选择部的处理例的示意图。

图 28 是表示同上的函数选择部的处理例的示意图。

图 29 是表示同上的音质变换装置的动作的流程图。

图 30 是表示同上的变形例 1 的音质变换装置的结构的结构图。

图 31 是表示同上的变形例 3 的音质变换装置的结构的结构图。

具体实施方式

下面，参照附图说明本发明的实施方式。

（实施方式 1）

图 4 是表示本发明的第 1 实施方式中的声音合成装置的结构的结构图。

本实施方式的声音合成装置可适当变换音质，包括：韵律推定部 101、单元存储部 102、选择部 103、函数存储部 104、拟合优度判断部 105、音质变换部 106、音质指定部 107、波形合成部 108。

单元存储部 102 作为单元存储机构构成，保存表示多种声音单元的信息。该声音单元根据预先收录的声音，按音素、音节、莫勒等单位进行保存。再有，单元存储部 102 也可以将声音单元作为声音波形或分析参数来保存。

函数存储部 104 作为函数保存机构构成，保存用于对保存在单元存储部 102 种的声音单元进行音质变换的多个变换函数。

这些多个变换函数与通过该变换函数可变换的音质相关联。例如，变换函数与表示“生气”、“高兴”、“悲伤”等感情的音质相关联。此外，变换函数例如与表示“DJ 风格”、“播音员风格”等讲话风格等的音质相关联。

变换函数的使用单位例如是声音单元、音素、音节、莫勒、重音句等。

例如使用共振峰频率的变形率或差分值、功率的变形率或差分值、基本频率的变形率或差分值等来生成变换函数。此外，变换函数也可以是将共振峰、功率或基本频率等分别同时变更的函数。

此外，变换函数中设定有可应用该函数的声音单元的范围。例如，

被设定为：若对预定的声音单元应用变换函数，则其使用结果被学习，从而该预定的声音单元被包含到变换函数的应用范围内。

此外，通过对表示“生气”等感情的音质的变换函数改变变量，来对音质进行内插，能够实现连续的音质变换。

韵律推定部 101 作为生成机构来构成，取得例如基于用户操作生成的文本数据。之后，韵律推定部 101 根据表示该文本数据中包含的各音素的音素信息，来对每个音素推定音韵环境、基本频率、持续时间长度，功率等韵律特征（韵律），并生成音素和表示该韵律的韵律信息。该韵律信息作为最终输出的合成声音的目标来使用。韵律推定部 101 向选择部 103 输出该韵律信息。此外，除音素信息之外，韵律推定部 101 也可以取得词素信息、重音信息、语法信息。

拟合优度判断部 105 作为类似度导出机构构成，判断存储在单元存储部 102 中的声音单元和存储在函数存储部 104 中的变换函数之间的拟合优度。

音质指定部 107 作为音质指定机构而构成，取得由用户指定的合成声音的音质，并输出表示其音质的音质信息。该音质表示例如“生气”、“高兴”、“悲伤”等感情或“DI 风格”、“播音员风格”等讲话风格等。

选择部 103 作为选择机构而构成，根据从韵律推定部 101 输出的韵律信息、从音质指定部 107 输出的音质、以及由拟合优度 105 判断的拟合优度，从单元存储部 102 选择最佳的声音单元，并且，从函数存储部 104 选择最佳的变换函数。即，选择部 103 根据拟合优度来相辅地选择声音单元和变换函数。

音质变换部 106 作为使用机构而构成，对于由选择部 103 选择的声音单元使用由选择部 103 选择的变换函数。即，音质变换部 106 通过用该变换函数变换声音单元，来生成由音质指定部 107 指定的音质的声音单元。本实施方式中，由该音质变换部 106 和选择部 103 构成了变换机构。

波形合成部 108 根据由音质变换部 106 变换的声音单元生成并输出

声音波形。例如，波形合成部 108 通过波形连接型声音合成方法、分析合成型声音合成方法，来生成声音波形。

在上述的声音合成装置中，当文本数据所包含的音质信息表示一连串的音素和韵律时，选择部 103 从单元存储部 102 选择与该音素信息对应的一连串声音单元（声音单元系列），并从函数存储部 104 选择与该音素信息对应的一连串的变换函数（变换函数系列）。之后，音质变换部 106 分别处理由选择部 103 选择的声音单元系列及变换函数系列的各自中包含的声音单元和变换函数。此外，波形合成部 108 根据由音质变换部 106 变换了的一连串声音单元，生成并输出声音波形。

图 5 是表示选择部 103 的结构的结构图。

选择部 103 具有单元点阵确定部 201、函数点阵确定部 202、单元成本判断部 203、成本综合部 204 以及检索部 205。

单元点阵确定部 201 根据从韵律推定部 101 输出的韵律信息，从存储在单元存储部 102 中的多个声音单元中确定最终应选择的声音单元的多个候补。

例如，单元点阵确定部 201 将所有的表示与韵律信息中包含的音素相同的音素的声音单元确定为候补。此外，单元点阵确定部 201 将韵律信息中包含的音素和韵律的类似度成为规定的阈值以内（例如，基本频率的差分在 20Hz 以内的情况等）的声音单元确定为候补。

函数点阵确定部 202 根据韵律信息、从音质指定部 107 输出的音质信息，从存储在函数存储部 104 中的多个变换函数中确定最终应选择的变换函数的几个候补。

例如，函数点阵确定部 202 将包含在韵律信息中的音素作为应用对象，将可变换为由音质信息表示的音质（例如“生气”的音质）的变换函数作为候补。

单元成本判断部 203 判断由单元点阵确定部 201 确定的声音单元候补和韵律信息的单元成本。

例如，单元成本判断部 203 将连接了由韵律推定部 101 推定的韵律

和声音单元候补的韵律的类似度、及声音单元时的连接边界附近的平滑程度用作最近似度，来判断单元成本。

成本综合部 204 综合由拟合优度判断部 105 判断的拟合优度和由单元成本判断部 203 判断的单元成本。

检索部 205 从由单元点阵确定部 201 确定的声音单元候补、由函数点阵确定部 202 确定的变换函数候补中，选择由成本综合部 204 计算的成本的价值成为最小的声音单元和变换函数。

下面，对选择部 103 和拟合优度判断部 105 进行具体说明。

图 6 是用于说明单元点阵确定部 201 和函数点阵确定部 202 的动作的说明图。

例如，韵律推定部 101 取得表示“红”的文本数据（单元信息），并输出包含在该音素信息中的包括各音素和各韵律的韵律信息组 11。该韵律信息组 11 包括：音素 a 和表示与其对应的韵律的韵律信息 t_1、音素 k 和表示与其对应的韵律的韵律信息 t_2、音素 a 和表示与其对应的韵律的韵律信息 t_3、音素 i 和表示与其对应的韵律的韵律信息 t_4。

单元点阵确定部 201 取得该韵律信息组 11，来确定声音单元候补组 12。该声音单元候补组 12 包括：对音素 a 的声音单元候补 u_{11}、u_{12}、u_{13}，对音素 k 的声音单元候补 u_{21}、u_{22}，对音素 a 的声音单元候补 u_{31}、u_{32}、u_{33}，对音素 i 的声音单元候补 u_{41}、u_{42}、u_{43}、u_{44}。

函数点阵确定部 202 取得上述韵律信息组 11 和音质信息，来确定例如与“生气”的音质对应的变换函数候补组 13。该变换函数候补组 13 包括：对音素 a 的变换函数候补 f_{11}、f_{12}、f_{13}，对音素 k 的变换函数候补 f_{21}、f_{22}、f_{23}，对音素 a 的变换函数候补 f_{31}、f_{32}、f_{33}、f_{34}、对音素 i 的变换函数候补 f_{41}、f_{42}。

单元成本判断部 203 计算表示由单元点阵确定部 201 确定的声音单元候补的最近似程度的单元成本 ucost（t_i,u_{ij}）。该单元 ucost（t_i,u_{ij}）是根据由韵律推定部 101 推定的音素所应具有的韵律信息韵律信息 t_i 和声音单元候补 u_{ij} 的类似度来判断的成本。

在此，韵律信息 t_i 表示对由韵律推定部 101 推定的音素信息的第 i 个音素的音韵环境、基本频率、持续时间长度和功率等。此外，声音单元候补 u_{ij} 是对第 i 个音素的第 j 个声音单元候补。

例如，单元成本判断部 203 计算综合了音韵环境的一致度、基本频率的误差、持续时间长度的误差、功率的误差和连接了声音单元时的连接变形等的单元成本。

拟合优度判断部 105 计算声音单元候补 u_{ij} 和变换函数候补 f_{ij} 的拟合优度 fcost（u_{ij},f_{ik}）。在此，变换函数候补 f_{jk} 是对第 i 个音素的第 k 个变换函数候补。由公式 1 定义该拟合优度 fcost（u_{ii},f_{ik}）。

公式 1：

$$f\cos t(u_{ij}, f_{ik}) = static\_\cos t(u_{ij}, f_{ik}) + dynamic\_\cos t(u_{(i-1)j}, u_{ij}, u_{(i+1)j}, f_{ik}) \cdots (式1)$$

在此，static#cost（u_{ij},f_{ik}）是声音单元候补 u_{ij}、（声音单元候补 u_{ij} 的音响特征）和变换函数候补 f_{ik}（在制作变换函数候补 f_{ik} 时使用的声音单元的音响特征）的静态拟合优度（类似度）。这样的静态拟合优度通过例如在制作变换函数候补时使用的声音单元的音响特征、即假定为可适当使用变换函数的音响特征（例如，共振峰频率、基本频率、功率、倒频谱系数（cepstral coefficients）等）与声音单元候补的音响特征的类似度来表现。

另外，静态拟合优度不限于这些，只要利用声音单元和变换函数中的某个的类似度就可以。此外，对于所有的声音单元和变换函数，当预先在未连线的状态下计算静态拟合优度，并对各声音单元使拟合优度对应上位的变换函数，计算静态拟合优度时，可以只将与该声音单元对应的变换函数设定为对象。

另一方面，dynamic#cost（$u_{(i-1)j}$, u_{ij}, $u_{(i+1)j}$, f_{ik}）是动态拟合优度，是对象的变换函数候补 f_{ik} 和声音单元候补 u_{ij} 的前后环境之间的拟合优度。

图 7 是用于说明动态拟合优度的说明图。

动态拟合优度例如根据学习数据来计算。

变换函数是根据通常发音的声音单元与基于感情或讲话风格来学习发音的声音单元的差分值来学习（制作）的。

例如图 7 的（b）所示，学习数据表示对一连串的声音单元候补（系列）u_{11}、u_{12}、u_{13} 中的声音单元候补 u_{12}，提高了基本频率 F_0 的变换函数 F_{12} 所学习的情况。此外，如图 7（c）所示，学习数据表示对一连串的声音单元候补（系列）u_{21}、u_{22}、u_{23} 中的声音单元候补 u_{22}，提高了基本频率 F_0 的变换函数 F_{22} 所学习的情况。

拟合优度判断部 105 在对图 7（a）所示的声音单元候补 u_{32} 选择变换函数时，根据包含 u_{32} 的前后声音单元的环境（u_{31}，u_{32}，u_{33}）、和变换函数候补（f_{11}，f_{22}）的学习数据环境（u_{11}，u_{12}，u_{13} 和 u_{21}，u_{22}，u_{23}）的一致度，来判断拟合优度。

在图 7 所示的情况下，（a）的学习数据所表示的环境是基本频率 F_0 随时间 t 而增加的环境，因此，如图（c）的学习数据所示，拟合优度判断部 105 判断为在基本频率 F_0 增加的环境下学习（生成）的变换函数 f_{22} 的动态拟合优度高（dynamic#cost 的值小）。

即，由于图 7（a）所示的声音单元候补 u_{33} 是基本频率 F_0 与时间 t 一起增加的环境，因此，如图（b）所示，拟合优度判断部 105 将在图 7（b）所示的基本频率 F_0 减少的环境中学习了的变换函数 f_{12} 的动态拟合优度计算为较低，将在图 7（c）所示的基本频率 F_0 增加的环境中学习了的变换函数 f_{22} 的动态拟合优度计算为较高。

换言之，拟合优度判断部 105 判断为：同要抑制前后环境的基本频率 F_0 相比，要进一步促进前后环境的基本频率 F_0 的增加的变换函数 f_{22} 的与图 7（a）所示前后环境的拟合优度更高。即，拟合优度判断部 105 判断为，对于声音单元候补 u_{32} 应选择变换函数候补 f_{22}。反之，若选择变换函数 f_{12}，则不能将具有变换函数 f_{22} 的变换特性反映到声音单元候补 u_{32}。此外，可以说，动态拟合优度是应该应用变换函数候补 f_{ik} 的一连串声音单元（在制作变换函数候补 f_{ik} 时使用的一连串声音单元）的动态特性与一连串声音单元候补 u_{ij} 的动态特性之间的类似度。

再有，图 7 中使用了基本频率的 F_0 动态特性，但本发明并不限于此，例如，也可以使用功率、持续时间长度、共振峰频率、倒频谱系数等。此外，不限于上述功率等的单个，而可以组合基本频率、功率、持续时间长度、共振峰频率、倒频谱系数等来计算动态拟合优度。

成本综合部 204 计算综合成本 manage#cost（t_i，u_{ij}，f_{ik}）。由公式 2 定义该综合成本。

公式 2：

$$manage\_cost(t_i,u_{ij},f_{ik}) = u\,cost(t_i,u_{ij}) + f\,cost(u_{ij},f_{ik}) \cdots \text{（式2）}$$

此外，在公式 2 中，分别将单元成本 ucost（t_i，u_{ij}）和拟合优度发 fcost（t_{ij}，f_{ik}）均等地相加，但也可以分别附以权重之后相加。

检索部 205 从由单元点阵确定部 201 和函数点阵确定部 202 确定的声音单元候补和变换函数候补中，选择由成本综合部 204 计算的综合成本的累加值成为最小的声音单元系列 U 和变换函数系列 F。例如，如图 6 所示，检索部 205 选择声音单元系列 U（u_{11}，u_{21}，u_{31}，u_{44}）和变换函数系列 F（f_{13}，f_{22}，f_{32}，f_{41}）。

具体来说，检索部 205 根据公式 3 选择上述的声音单元系列 U 和变换函数系列 F。再有，n 表示音素信息中所包含的音素的个数。

公式 3：

$$U,F = \underset{u,f}{\arg\min} \sum_{i=1,2,\ldots,n} manage\_cost(t_i,u_{ij},f_{ik}) \cdots \text{（式3）}$$

图 8 是表示上述选择部 103 的动作的流程图。

首先，选择部 103 确定几个声音单元候补和变换函数候补（步骤 S100）。接着，选择部 103 对 n 个韵律信息 t_i、对应于各韵律信息 t_i 的 n' 个声音单元候补和对应于各韵律信息 t_i 的 n"个变换函数候补的各组合，计算综合成本 manage#cost（t_i，u_{ij}，f_{ik}）（步骤 S102～S106）。

选择部 103 为了计算综合成本，首先计算单元成本 ucost(t_i，u_{ij})（步骤 S102），并计算拟合优度 fcost(u_{ij} ，f_{ik})。此外，选择部 103 通过将在步

骤 S102、S104 算出的单元成本 ucost(t_i，u_{ij})和拟合优度 fcost(u_{ij} ，f_{ik})相加，来计算综合成本 manage#cost（t_i，u_{ij}，f_{ik}）。这样的综合成本的计算，是通过选择部 103 的检索部 205 对单元成本判断部 203 和拟合优度判断部 105 指示改变 i、j、k，来对各 i、j、k 的各组合进行。

接着，选择部 103 在个数 n'、n"的范围内改变 j、k 来累加 i=1~n 的各综合成本 manage#cost（t_i，u_{ij}，f_{ik}）（步骤 S108）。之后，选择部 103 选择该累加值成为最小的声音单元系列 U 和变换函数系列 F（步骤 S110）。

此外，图 8 中，预先计算成本值之后，选择了累加值成为最小的声音单元系列 U 和变换函数系列 F，但也可以使用检索问题中所使用的 Viterbi 算法来选择声音单元系列 U 和变换函数系列 F。

图 9 是表示本实施方式的声音合成装置的动作的流程图。

声音合成装置的韵律推定部 101 取得包含音素信息的文本数据，并根据该音素信息来推定各音素应具有的基本频率、持续时间长度、功率等韵律性特征（韵律）（步骤 S200）。例如，韵律推定部 101 通过使用了数量化 1 类的方法来进行推定。

之后，声音合成装置的音质指定部 107 取得用户所指定的合成声音的音质，例如“生气”的音质（步骤 S202）。

声音合成装置的选择部 103 根据表示韵律推定部 101 的推定结果的韵律信息和由音质指定部 107 取得的音质，从单元存储部 102 确定声音单元候补（步骤 S204），并且，从函数存储部 104 确定表示“生气”的变换函数候补（步骤 S206）。之后，选择部 103 从被确定的声音单元候补和变换函数候补选择综合成本成为最小的声音单元和变换函数（步骤 S208）。即，在音素信息表示一连串的音素的情况下，选择部 103 选择综合成本的累加值成为最小的声音单元系列 U 和变换函数系列 F。

接着，声音合成装置的音质变换部 106 使用变换函数系列 F，对在步骤 S208 被选择的声音单元系列 U 进行音质变换（步骤 S210）。声音合成装置的波形合成部 108 根据被音质变换部 106 进行了音质变换的声音单元系列 U，生成并输出声音波形（步骤 S212）。

如上所述，在本实施方式中，对每个声音单元应用最佳的变换函数，因此，能够适当地变换音质。

再此，将本实施方式与现有技术（特开 2002-215198 号公报）进行比较，来详细说明本实施方式的效果。

上述现有技术的声音合成装置，按元音和辅音等的各种类型制作频谱包络变换表（变换函数），对属于某种类型的声音单元，应用设定在该类型中的频谱包络变换表。

但是，若将由类型代表的频谱包络变换表应用于类型中的所有声音单元，则产生例如如下问题：在变换后的声音中多个共振峰频率过于接近，或者，变换后的声音的频率超过奈奎斯特频率。

具体地，用图 10 和图 11 说明上述问题。

图 10 是表示元音“i”的声音频谱的图。

图 10 中的 A101、A102、A103 表示频谱强度高的部分（频谱的峰值）。

图 11 是表示元音“i”以外的其他声音的频谱的图。

与图 10 同样，图 11 中的 B101、B102、B103 表示频谱强度高的部分。

如上述的图 10 和图 11 所示，即使是相同的元音“i”，有时频谱的形状也大不相同。因此，在以代表类型的声音（声音单元）为基础制作频谱包络变换表的情况下，若对与代表声音单元的频谱大不相同的声音单元使用该频谱包络变换表，则有时不能得到预想的音质变换效果。

用图 12A 和图 12B 说明更具体的例子。

图 12A 是表示对元音“i”的频谱应用变换函数的例子的图。

变换函数 A202 是对图 10 所示的元音“i”的声音制作的频谱包络变换表。频谱 A201 表示代表类型的声音单元（例如图 10 所示的元音“i”）的频谱。

例如，若对频谱 A201 使用变换函数 A202，则频谱 A201 变换为频谱 A203。该变换函数 A202 对中间频带频率进行了提升到高频带的变换。

但是，如图 10 和 11 所示，即使两个声音单元是相同的元音“i”,它

们的频谱有时也大不相同。

图 12B 是表示对元音“i”的其它频谱应用了变换函数的例子的图。

频谱 B201 是例如图 11 所示的元音“i”的频谱，与图 12A 的频谱 A201 大不相同。

若对该频谱 201 应用变换函数 A202，则频谱 B102 变换为频谱 B203。即，频谱 B203 中，该频谱的第 2 峰值和第 3 峰值显著接近，形成一个峰值。这样，若对频谱 B201 应用变换函数 A202，则不能得到与对频谱 A201 应用了变换函数 A202 时的音质变换同样的音质变换效果。此外，在上述现有技术中，存在有如下的问题：在变换后的频谱 B203 中两个峰值过于接近而形成一个峰值，损害元音“i”的音韵性。

另一方面，在本发明的实施方式的声音合成装置中，将声音单元的音响特征和作为变换函数的源数据的声音单元的音响特征，并将两个声音单元的音响特征最接近的声音单元和变换函数对应起来。接着，本发明的声音合成装置对声音单元的音质利用与该声音单元对应的变换函数来进行变换。

即，本发明的声音合成装置保持多个对元音“i”的变换函数候补，并根据在制作变换函数时使用的声音单元的音响特征，来选择对作为变换对象的声音单元最佳的变换函数，将该选择的变换函数应用于声音单元。

图 13 是用于说明本实施方式的声音合成装置适当地选择变换函数的情况的说明图。再有，图 13（a）示出变换函数（变换函数候补）n、和在制作该变换函数候补 n 时使用了的声音单元的音响特征；图 13（b）表示变换函数（变换函数候补）m、和在制作该变换函数候补 m 时使用了的声音单元的音响特征。此外，图 13（c）表示变换对象的声音单元的音响特征。在此，（a）、（b）和（c）中，利用第 1 共振峰 F1、第 2 共振峰 F2、第 3 共振峰 F3 来用图表表示音响特征，该图表的横轴表示时间，该图表的纵轴表示频率。

本实施方式中的声音合成装置例如从（a）所示的变换函数候补 n 和

（b）所示的变换函数候补 m 中，将音响特征与（c）所示的变换对象的声音单元类似的变换函数候补作为变换函数选择。

在此，（a）所示的变换函数候补 n 进行使第 2 共振峰 F2 降低 100Hz 的变换、使第 3 共振峰 F3 降低 100Hz 的变换。另一方面，（b）所示的变换函数候补 m 进行将第 2 共振峰 F2 提高 500Hz、将第 3 共振峰 F3 降低 500Hz。

这样的情况下，本实施方式的声音合成装置计算（c）所示的变换对象的声音单元的音响特征、和在制作（a）所示的变换函数候补 n 时所使用的声音单元的音响特征之间的类似度，并计算（c）所示的变换对象的声音单元的音响特征、和在制作（b）所示的变换函数候补 m 时所使用的声音单元的音响特征之间的类似度。其结果，本实施方式中的声音合成装置在第 2 共振峰 F2 和第 3 共振峰 F3 的频率中，能够判断为变换函数候补 n 的音响特征与变换函数候补 m 的音响特征相比，与变换函数候补 n 的音响特征更类似。因此，声音合成装置将变换函数候补 n 作为变换函数选择，并将该变换函数 n 应用于变换对象的声音单元。这时，声音合成装置利用各共振峰的移动量来进行频谱包络的变形。

在此，如上述现有技术的声音合成装置，在使用类型代表函数（例如，图 13（b）所示的变换函数候补 m）的情况下，第 2 共振峰和第 3 共振峰交叉，从而不仅得不到音质变换效果，还不能确保音韵性。

而在本发明的声音合成装置中，通过利用类似度（拟合优度）来选择变换函数，对图 13（c）所示的变换对象的声音单元使用以与该声音单元的音响特征接近的声音单元为基础制作的变换函数。因此，在本实施方式中，在变换后的声音中，能够消除共振峰频率分别过于接近、或该声音的频率超过奈奎斯特频率的问题。此外，在本实施方式中，对于作为变换函数制作源的声音单元（例如，具有图 13（a）所示的音响特征的声音单元）类似的声音单元（例如，具有图 13（c）所示的音响特征的声音单元）应用该变换函数，因此，能够得到与将该变换函数应用于制作源的声音单元时所得到的音质变换效果相同的效果。

如上所述，在本实施方式中，不像上述现有的声音合成装置那样，不被声音单元的类型等而左右，而能够对各声音单元分别选择最适合的变换函数，能够将音质变换的变形抑制在最小限度上。

此外，在本实施方式中，由于用变换函数变换音质，能够连续变换音质，并且能够生成数据库（单元存储部 102）中所没有的音质的声音波形。此外，在本实施方式中，由于如上所述能够对每个声音单元使用最佳的变换函数，因此，不用进行无用的校正即可将声音波形的共振峰频率抑制在适当的范围内。

此外，在本实施方式中，从单元存储部 102 和函数存储部 104 同时相辅地选择文本数据和用于实现由音质指定部 107 指定的音质的声音单元和变换函数。即，在找不到与声音单元对应的变换函数的情况下，变更为不同的声音单元。此外，在找不到与变换函数对应的声音单元的情况下，变更为不同的变换函数。由此，能够同时对与该文本数据对应的合成声音的质量和变换为由音质指定部 107 指定的音质的质量进行最优化，能够得到高音质（质量）且所希望的音质的合成声音。

再有，在本实施方式中，选择部 103 根据综合成本的结果来选择了声音单元和变换函数，但也可以选择由拟合优度判断部 105 计算的静态拟合优度、动态拟合优度或者将这些组合的拟合优度成为规定的阈值以上的声音单元和变换函数。

（变形例）

上述实施方式 1 的声音合成装置根据指定的一个音质，来选择声音单元系列 U 和变换函数系列 F（声音单元和变换函数）。

本变形例的声音合成装置接受多个音质的指定，并根据该多个音质来选择声音单元系列 U 和变换函数系列 F。

图 14 是用于说明本变形例的单元点阵确定部 201 和函数点阵确定部 202 的动作的说明图。

函数点阵确定部 202 确定用于实现由函数存储部 104 指定的多个音质的变换函数候补。例如，在由音质指定部 107 接受了“生气”和“高

兴”的音质的指定的情况下，函数点阵确定部 202 从函数存储部 104 确定与“生气”和“高兴”的各音质对应的变换函数候补。

例如，如图 14 所示，函数点阵确定部 202 确定变换函数候补组 13。该变换函数候补组 13 中包含与“生气”的音质对应的变换函数候补组 14 和与“高兴”的音质对应的变换函数候补组 15。变换函数候补组 14 包括：对应于音素 a 的变换函数候补 f_{11}，f_{12}，f_{13}、对应于音素 k 的变换函数候补 f_{21}，f_{22}，f_{23}、对应于音素 a 的变换函数候补 f_{31}，f_{32}，f_{33}，f_{34}、对应于音素 i 的变换函数候补 f_{41}，f_{42}。变换函数候补组 15 包括：对应于音素 a 的变换函数候补 g_{11}，g_{12}、对应于音素 k 的变换函数候补 g_{21}，g_{22}，g_{23}、对应于音素 a 的变换函数候补 g_{31}，g_{32}，g_{33}、对应于音素 i 的变换函数候补 g_{41}，g_{42}，g_{43}。

拟合优度判断部 105 计算声音单元候补 u_{ij}、变换函数候补 f_{ik} 和变换函数候补 g_{ih} 之间的拟合优度 fcost（u_{ij}，f_{ik}，g_{ih}）。在此，变换函数候补是对第 i 个音素的第 h 个变换函数候补。

根据公式 4 计算该拟合优度 fcost（u_{ij}，f_{ik}，g_{ih}）。

公式 4

$$f\cos t(u_{ij},f_{ik},g_{ih})=f\cos t(u_{ij},f_{ik})+f\cos t(u_{ij}*f_{ik},g_{ih})\cdots(式4)$$

在此，公式 4 中所示的 $u_{ij}*f_{ik}$ 表示对单元使用了变换函数之后的声音单元。

成本综合部 204 使用单元选择成本 ucost（t_i，u_{ij}）和拟合优度 fcost（u_{ij}，f_{ik}，g_{ih}），来计算综合成本 manage#cost（t_i，u_{ij}，f_{ik}，g_{ih}）。根据公式 5 计算该综合成本 manage#cost（t_i，u_{ij}，f_{ik}，g_{ih}）。

公式 5：

$$manage\_\cos t(t_i,u_{ij},f_{ik},g_{ih})=u\cos t(t_i,u_{ij})+f\cos t(u_{ij},f_{ik},g_{ih})\cdots(式5)$$

检索部 205 根据公式 6 选择声音单元系列 U 和变换函数系列 F、G。

公式 6：

$$U,F,G=\underset{u,f,g\ i=1,2,...,n}{\arg\min}\sum manage\_\cos t(t_i,u_{ij},f_{ik},g_{ih}) \cdots \text{(式6)}$$

例如，如图 14 所示，选择部 103 选择声音单元系列 U（u_{11}，u_{21}，u_{32}，u_{44}）、变换函数系列 F（f_{13}，f_{22}，f_{32}，f_4，）和变换函数系列 G（g_{12}，g_{22}，g_{32}，g_{41}）。

如上所述，在本变形例中，音质指定部 107 接受多个音质的指定，来计算基于这些音质的拟合优度和综合成本，因此，能够同时对与文本数据对应的合成声音的质量和向上述多个音质的变换的质量进行最优化。

再有，在本实施方式中，拟合优度判断部 105 在拟合优度 fcost（u_{ij}，f_{ik}）上加上拟合优度 fcost（u_{ij}*f_{ik}，g_{ih}），来计算最终的拟合优度 fcost（u_{ij}，f_{ik}，g_{ih}），但是也可以拟合优度 fcost（u_{ij}，f_{ik}）上加上拟合优度 fcost（u_{ij}，g_{ih}），来计算最终的拟合优度 fcost（u_{ij}，f_{ik}，g_{ih}）。

此外，在本实施例中，音质指定部 107 接受了两个音质的指定，但是也可以接受 3 个以上的音质的指定。在这样的情况下，本变形例中，拟合优度判断部 105 用与上述同样的方法计算拟合优度，并将与各音质对应的变换函数应用于声音单元。

（实施方式 2）

图 15 是表示本发明实施方式 2 的声音合成装置结构的结构图。

本实施方式的声音合成装置包括：韵律推定部 101、单元存储部 102、单元选择部 303、函数存储部 104、拟合优度判断部 302、音质变换部 106、音质指定部 107、函数选择部 301、波形合成部 108。再有，本实施方式的构成要素中，对于与实施方式 1 的声音合成装置的构成要素相同的构件，标注了与实施方式 1 的构成要素相同的标记，并省略详细说明。

在此，在本实施方式的声音合成装置中，首先，函数选择部 301 根据由音质指定部 107 指定的音质和韵律信息来选择变换函数（变换函数系列），并由单元选择部 303 根据该变换函数选择声音单元（声音单元系列），这一点与实施方式 1 不同。

函数选择部 301 作为函数选择机构构成，根据从韵律推定部 101 输出的韵律信息和从音质指定部 107 输出的音质信息，从函数存储部 104 选择变换函数。

单元选择部 303 作为单元选择机构而构成，根据从韵律推定部 101 输出的韵律信息，从单元存储部 102 确定几个声音单元的候补。并且，单元选择部 303 从该候补中选择与该韵律信息和由函数选择部 301 选择的变换函数最合适的声音单元。

拟合优度判断部 302 利用与实施方式 1 的拟合优度判断部 105 相同的方法，来判断由函数选择部 301 已选择的变换函数和由单元选择部 303 确定的几个声音单元候补之间的拟合优度 fcost（u_{ij}，f_{ik}）。

音质变换部 106 对由单元选择部 303 选择的声音单元，应用由函数选择部 301 选择的变换函数。由此，音质变换部 106 生成由用户在音质指定部 107 指定的音质的声音单元。本实施方式中，由该音质变换部 106、函数选择部 301 和单元选择部 303 构成变换机构。

波形合成部 108 根据由音质变换部 106 变换的声音单元生成并输出声音波形。

图 16 是表示函数选择部 301 的结构的结构图。

函数选择部 301 包括函数点阵确定部 311 和检索部 312。

函数点阵确定部 311 从存储在函数存储部 104 中的变换函数中，将几个变换函数确定为用于变换为由音质信息表示的音质（被指定的音质）的变换函数候补。

例如，在音质指定部 107 接受了“生气”的音质的指定的情况下，函数点阵确定部 311 从函数存储部 104 中存储的变换函数中，把用于变换为“生气”的音质的变换函数确定为候补。

检索部 312 从由函数点阵确定部 311 确定的几个变换函数候补中，选择对从韵律推定部 107 输出的韵律信息适当的变换函数。例如，韵律信息包括音素系列、基本频率、持续时间长度和功率等。

具体而言，检索部 312 选择一连串韵律信息 t_i 和一连串变换函数候补

f_{ik} 的拟合优度（在学习变换函数候补 f_{ik} 时所使用的声音单元的韵律特征和韵律信息 t_i 的类似度）最大、即如满足公式 7 的满足一连串变换函数的变换函数系列 F（f_{1k}，f_{2k}，…，f_{nk}）。

公式 7：

$$F = \arg\min_{f} \sum_{i=1,\cdots,n} f\,\mathrm{cost}(t_i, f_{ik}) = static\_\mathrm{cost}(t_i, f_{ik}) + dynamic\_\mathrm{cost}(t_{i-1}, t_i, t_{i+1}, f_{ik}) \cdots \text{（式7）}$$

在此，本实施方式中，如图 7 所示，在计算拟合优度时所使用的项目只是基本频率、持续时间长度、功率等韵律信息 t_i，这一点与实施方式 1 的公式 1 所表示的拟合优度不同。

此外，检索部 312 将所选择的候补作为用于变换为被指定的音质的变换函数（变换函数系列）来输出。

图 17 是表示单元选择部 303 结构的结构图。

单元选择部 303 具备单元点阵确定部 321、单元成本判断部 323、成本综合部 324、检索部 325。

这样的单元选择部 303 选择从韵律推定部 101 输出的韵律信息和最符合从函数选择部 301 输出的变换函数的声音单元。

单元点阵确定部 321 与实施方式 1 的单元点阵确定部 321 同样，根据由韵律推定部 101 输出的韵律信息，从单元存储部 102 中存储的多个声音单元中确定几个声音单元候补。

单元成本判断部 323 与实施方式 1 的单元成本判断部 203 同样，判断由单元点阵确定部 321 确定的声音单元候补和韵律信息的单元成本。即，单元成本判断部 323 计算由单元点阵确定部 321 确定的声音单元候补的最近似程度的单元成本 ucost（t_i，u_{ij}）。

成本综合部 324 与实施方式 1 的成本综合部 204 同样，通过综合由拟合优度判断部 302 判断的拟合优度和由单元成本判断部 323 判断的单元成本，计算综合成本 manage#cost（t_i，u_{ij}，f_{ik}）。

检索部 325 从由单元点阵确定部 321 确定的声音单元候补中，选择由成本综合部 324 计算出的综合成本的累加值成为最小的声音单元系列

U。

具体来说，检索部 325 根据公式 8 来选择上述的声音单元系列 U。

公式 8：

$$U = \arg\min_{u} \sum_{i=1,2,...,n} manage\_cost(t_i, u_{ij}, f_{ik}) \quad \cdots \quad (式8)$$

图 18 是表示本实施方式中的声音合成装置的结构的流程图。

声音合成装置的韵律推定部 101 取得包含音素信息的文本数据，并根据该音素信息，来推定各音素所应具有的基本频率、持续时间长度、功率等韵律性特征（韵律）（步骤 S300）。例如，韵律推定部 101 利用采用了数量化 I 类的方法来进行推定。

接着，声音合成装置的音质指定部 107 取得用户所指定的合成声音的音质例如“生气”的音质（步骤 S302）。

声音合成装置的函数选择部 301 根据被音质指定部 107 取得的音质，从函数存储部 104 中确定表示“生气”的音质的变换函数候补（步骤 S304）。之后，函数选择部 301 从该变换函数候补中选择与表示韵律推定部 101 的推定结果的韵律次信息最合适的变换函数（步骤 S306）。

声音合成装置的单元选择部 303 根据韵律信息，从单元存储部 102 确定几个声音单元的候补（步骤 S308）。此外，单元选择部 303 从该候补中选择与该韵律信息以及由函数选择部 301 选择的变换函数最适合的声音单元（步骤 S310）。

接着，声音合成装置的音质变换部 106 将在步骤 S306 选择的变换函数应用于在步骤 S310 被选择的声音单元，进行音质变换（步骤 S312）。声音合成装置的波形合成部 108 根据由音质变换部 106 进行了音质变换的声音单元，生成并输出声音波形（步骤 S314）。

在上述的本实施方式中，首先，根据音质信息和韵律信息选择变换函数，并选择对该选择的变换函数最佳的声音单元。作为该实施方式的较佳状况，有时不能充分确保变换函数。具体而言，在准备对各种音质的变换函数时，对各音质准备多个变换函数是较困难的。在这样的情况

下，即使函数存储部 104 中存储的变换函数的个数少，只要是单元存储部 102 中存储的声音单元的个数充分多，则能够同时最优化与文本数据对应的合成声音的质量和向由音质指定部 107 指定的音质变换的质量。

此外，与同时选择声音单元和变换函数的情况相比，能够减少计算量。

此外，在本实施方式中，单元选择部 303 根据综合成本的结果选择了声音单元，但也可以选择由拟合优度判断部 302 计算的静态拟合优度、动态拟合优度或组合它们的拟合优度大于等于预定的阈值的声音单元。

（实施方式 3）

图 19 是表示本发明的第 3 实施方式的声音合成装置结构的结构图。

本实施方式的声音合成装置包括：韵律推定部 101、单元存储部 102、单元选择部 403、函数存储部 104、拟合优度判断部 402、音质变换部 106、音质指定部 107、函数选择部 401、波形合成部 108。再有，本实施方式的构成要素中，对于与实施方式 1 的声音合成装置的构成要素相同的构件，标注与实施方式 1 的构成要素相同的标记，省略详细说明。

在此，在本实施方式的声音合成装置中，首先单元选择部 403 根据从韵律推定部 101 输出的韵律信息来选择声音单元（声音单元系列），并由函数选择部 401 根据该声音单元选择变换函数（变换函数系列），这一点与实施方式 1 不同。

单元选择部 403 从单元存储部 102 选择与从韵律推定部 101 输出的韵律信息最合适的声音单元。

函数选择部 401 根据音质信息和韵律信息，从函数存储部 104 确定几个变换函数的候补。此外，函数选择部 401 从该候补中选择适合由单元选择部 403 选择的声音单元的变换函数。

拟合优度判断部 402 通过与实施方式 1 的拟合优度判断部 105 相同的方法，判断已由单元选择部 403 选择的声音单元和由函数选择部 401 确定的几个变换函数候补之间的拟合优度 fcost（u_{ij}，f_{ik}）。

音质变换部 106 对由单元选择部 403 选择的声音单元，应用由函数

选择部 401 选择的变换函数。从而，音质变换部 106 生成由音质指定部 107 指定的音质的声音单元。

波形合成部 108 根据由音质变换部 106 变换了的声音单元生成并输出声音波形。

图 20 是表示单元选择部 403 的结构的结构图。

单元选择部 403 具备单元点阵确定部 411、单元成本判断部 412、检索部 413。

单元点阵确定部 411 与实施方式 1 的单元点阵确定部 201 同样，根据从韵律推定部 101 输出的韵律信息，从存储在单元存储部 102 中的多个声音单元中，确定几个声音单元候补。

单元成本判断部 412 与实施方式 1 的单元成本判断部 203 同样，判断由单元点阵确定部 411 确定的声音单元候补和韵律信息的单元成本。即，单元成本判断部 412 计算表示由单元点阵确定部 411 确定的声音单元候补的最近似程度的单元成本 ucost（t_i，u_{ij}）。

检索部 413 从由单元点阵确定部 411 确定的声音单元候补中，选择由单元成本判断部 412 计算的单元成本的累加值最小的声音单元系列 U。

具体而言，检索部 413 根据公式 9，选择上述的声音单元系列 U。

公式 9：

$$U = \arg\min_{u} \sum_{i=1,2,...,n} u\cos t(t_i, u_{ij}) \quad \text{（式 9）}$$

图 21 是表示函数选择部 401 的结构的结构图。

函数选择部 401 具备函数点阵确定部 421 和检索部 422。

函数点阵确定部 421 根据从音质指定部 107 输出的音质信息、从韵律推定部 101 输出的韵律信息，从函数存储部 104 确定几个变换函数候补。

检索部 422 从由函数点阵确定部 421 确定的几个变换函数候补中，选择最符合已由单元选择部 403 选择的声音单元的变换函数。

具体而言，检索部 422 根据公式 10，选择一连串的变换函数即变换

函数系列F（f_{1k}，f_{2k}，…，f_{nk}）。

公式10：

$$F = \arg\min_{f} \sum_{i=1,2,...,n} f\cos t(u_{ij}, f_{ik}) \quad \cdots \text{（式10）}$$

图22是表示本实施方式的声音合成装置的动作的流程图。

声音合成装置的韵律推定部101取得包含音素信息的文本数据，并根据该音素信息推定各音素所应具有的基本频率、持续时间长度、功率等韵律性特征（韵律）（步骤S400）。例如，韵律推定部101利用采用了数量化I类的方法来进行推定。

接着，声音合成装置的音质指定部107取得用户所指定的合成声音的音质例如“生气”的音质（步骤S402）。

声音合成装置的单元选择部403根据从韵律推定部101输出的韵律信息，从单元存储部102确定几个声音单元候补（步骤S404）。此外，单元选择部403从该声音单元候补中选择与该韵律信息最适合的声音单元（步骤S406）。

声音合成装置的函数选择部401根据音质信息和韵律信息，从函数存储部104中确定几个表示“生气”的音质的变换函数候补（步骤S408）。之后，函数选择部401从该变换函数候补中选择与表示由单元选择部403已选择的声音单元最合适的变换函数（步骤S410）。

接着，声音合成装置的音质变换部106将在步骤S410选择的变换函数应用于在步骤S406被选择的声音单元，进行音质变换（步骤S412）。声音合成装置的波形合成部108根据由音质变换部106进行了音质变换的声音单元，生成并输出声音波形（步骤S414）。

在上述的本实施方式中，首先，根据音质信息选择声音单元，选择对该被选择了的声音单元最佳的变换函数。作为该实施方式的较佳状况，例如，能确保足够变量的变换函数，但是有时不能确保足够变量的表示新讲话者的音质的声音单元。具体而言，一般即使将多个使用者的声音作为声音单元来使用，也很难收录大量的声音。在这样的情况下，即使

单元存储部 102 中存储的声音单元的个数少，如本实施方式那样，只要是函数存储部 104 中存储的变换函数的个数充分多，则能够同时最优化与文本数据对应的合成声音的质量和向由音质指定部 107 指定的音质变换的质量。

此外，与同时选择声音单元和变换函数的情况相比，能减少计算量。

此外，在本实施方式中，函数选择部 401 根据综合成本的结果选择了声音单元，但也可以选择由拟合优度判断部 402 计算的静态拟合优度、动态拟合优度或组合它们的拟合优度大于等于预定的阈值的声音单元。

（实施方式 4）

下面，用附图对本发明的第 4 实施方式进行详细说明。

图 23 是表示本发明实施方式的音质变换装置（声音合成装置）结构的结构图。

本实施方式的声音合成装置根据文本数据 501 生成表示音质 A 的声音的 A 声音数据 506，并将该音质 A 适当地变换为音质 B，其包括：文本分析部 502、韵律生成部 503、单元连接部 504、单元选择部 505、变换率指定部 507、函数应用部 509、A 单元数据库 510、A 基点数据库 511、B 基点数据库 512、函数提取部 513、变换函数数据库 514、函数选择部 515、第 1 缓冲器 517、第 2 缓冲器 518 和第 3 缓冲器 519。

此外，在本实施方式中，变换函数数据库 514 作为函数保存机构构成，函数选择部 515 作为类似度导出机构、代表值确定机构和选择机构来构成。此外，函数应用部 509 作为函数适用单元来构成。即，本实施方式中，由作为函数选择部 515 的选择机构的功能和作为函数应用部 509 的函数适用机构的功能来构成了变换机构。此外，文本分析部 502 作为分析机构构成，A 单元数据库 510 作为单元代表值存储机构构成，单元选择部 505 作为选择存储机构构成。再有，A 基点数据库 511 作为基准代表值存储机构构成，B 基点数据库 512 作为目标代表值存储机构构成，函数提取部 513 作为变换函数生成机构构成。此外，第 1 缓冲器 506 作为单元存储机构构成。

文本分析部502取得作为读取对象的文本数据501并进行语言分析，进行从假名和汉字交叉的文章向单元串（音素串）的变换或词素信息的提取等。

韵律生成部 503 根据该分析结果，生成包括附加在声音上的重音或各单元（音素）的持续时间长度等的韵律信息。

A 单元数据库 510 存储对应于音质 A 的声音的多个单元和附加在各单元上的表示该单元的音响特征的信息。以后，将该信息称作基点信息。

单元选择部 505 从 A 单元数据库 510 选择与所生成的语言分析结果和韵律信息对应的最佳单元。

单元连接部 504 通过连接被选择的单元，生成将文本数据 501 的内容作为音质 A 的声音表示的 A 声音数据 506。之后，单元连接部 504 将该 A 声音数据 506 存储到第 1 缓冲器 517 中。

A 声音数据 506 除了包含波形数据以外，还包含被使用的单元的基点信息和波形数据的标识信息。A 声音数据 506 中包含的基点信息是附加在单元选择部 505 所选择的各单元上的信息，标识信息是由单元连接部 504 根据韵律生成部 503 所生成的各单元的持续时间长度来生成的。

A 基点数据库 511 按照包含在音质 A 的声音中的各单元，存储着该单元的标识信息和基点信息。

B 基点数据库 512 对与 A 基点数据库 511 中的音质 A 的声音中包含的各单元对应的、包含在音质 B 的声音中的各个单元，存储着该单元的标识信息和基点信息。例如，如果 A 基点数据库 511 对音质 A 的声音“祝贺”中包含的各个单元存储着该单元的标识信息和基点信息，则 B 基点数据库 512 对音质 B 的声音“祝贺”中所包含的各个单元存储着该单元的标识信息和基点信息。

函数提取部 513 将分别与 A 基点数据库 511 和 B 基点数据库 512 对应的单元之间的标识信息及基点信息的差分，作为用于将各单元的音质从音质 A 变换为音质 B 的变换函数来生成。之后，函数提取部 513 将 A 基点数据库 511 的每个单元的标识信息及基点信息分别与如上述那样声

称的各单元的变换函数对应起来，存储到变换函数据库 514 中。

函数提取部 515 对 A 声音数据 506 中包含的每个单元部分，从变换函数数据库 514 选择与最接近该单元部分所具有的基点信息的基点信息对应的变换函数。从而，对 A 声音数据 506 中包含的各单元部分，能够自动高效地选择最适合于该单元部分的变换的变换函数。此外，函数选择部 515 将依次选择的所有变换函数作为变换函数数据 516 生成，并存储到第 3 缓冲器 519 中。

变换率指定部 507 对函数应用部 509 指定表示音质 A 的声音接近音质 B 的声音的比例的变换率。

函数应用部 509 用变换函数数据 516 将该 A 声音数据 506 变换为已变换声音数据 508，以使 A 声音数据 506 所表示的音质 A 的声音按由变换率指定部 507 指定的变换率接近音质 B 的声音。此外，函数应用部 509 将已变换声音数据 508 存储在第 2 缓冲器 518 中。这样被存储的已变换声音数据 508 被传递给声音输出用设备或记录用设备以及通信用设备等。

再有，本实施方式中，将声音的构成单位即单元（声音单元）作为音素进行了说明，但该单元也可以是其它构成单位。

图 24A 和图 24B 是表示本实施方式中的基点信息的一例的概略图。

基点信息是表示音素的基点的信息，下面，说明该基点。

如图 24A 所示，音质 A 的声音中包含的规定的音素部分的频谱中，表现了带有声音的音质的两个共振峰的轨迹 803。例如，该音素的基点 807 是作为两个共振峰的轨迹 803 所示的频率中的、与该音素的持续时间长度的中心 805 对应的频率定义。

和上述同样，如图 24B 所示，音质 B 的声音中包含的规定的音素部分的频谱中，表现了带有声音的音质的两个共振峰轨迹 804。例如，该音素的基点 808 是作为两个共振峰轨迹 804 所示的频率中的、与该音素的持续时间长度的中心 806 对应的频率定义。

例如，上述音质 A 的声音和上述音质 B 的声音在文章（内容）上相同，图 24A 所示的音素与图 24B 所示的音素对应的情况下，本实施方式

的音质变换装置利用上述基点 807、808，变换该音素的音质。即，本实施方式的音质变换装置对音质 A 的音素的声音频谱进行频率轴上的频谱伸缩，以使基点 807 表示的音质 A 的声音频谱的共振峰位置对准进入到由基点 808 表示的音质 B 的声音频谱的共振峰位置，而且，在时间轴上也进行伸缩，以使该音素的持续时间长度对准进入。由此，能够使音质 A 的声音与音质 B 的声音相似。

此外，在本实施方式中，将音素的中心位置的共振峰频率作为基点来定义，是因为元音的声音频谱在音素中心附近最稳定。

图 25A 和图 25B 是用于说明存储在 A 基点数据库 511 和 B 基点数据库 512 中的信息的说明图。

如图 25A 所示，A 基点数据库 511 中存储有包含在音质 A 的声音中的音素串和与该音素串的各音素对应的标识信息和基点信息。如图 25B 所示，B 基点数据库 512 中存储有包含在音质 B 的声音中的音素串和与该音素串的各音素对应的标识信息和基点信息。标识信息是表示声音中包含的各音素的讲话定时的信息，通过各音素的持续时间长度（持续长度）来表现。即，规定音素的讲话定时由到前一个音素为止的各音素的持续长度的总合来表示。此外，基点信息由用上述各音素的频谱表示的两个基点（基点 1 和基点 2）来表示。

例如，如图 25A 所示，A 基点数据库 511 中存储有音素串“ome”，并且，对于音素“o”，存储着持续时间长度（80ms）、基点 1（3000Hz）、基点 2（4300Hz）。此外，对于音素“m”，存储着持续长度（50ms）、基点 1（2500ms）、基点 2（4250Hz）。此外，音素“m”的讲话定时是，在从音素“o”开始讲话的情况下，是从该开始起经过了 80ms 的定时。

另一方面，如图 25B 所示，B 基点数据库 512 中存储着与上述 A 基点数据库对应的音素串“ome”，并且，对于音素“o”，存储着持续时间长度（70ms）、基点 1（3100Hz）、基点 2（4400Hz）。此外，对于音素“m”，存储着持续长度（40ms）、基点 1（2400ms）、基点 2（4200Hz）。

函数提取部 513 根据包含在 A 基点数据库 511 和 B 基点数据库 512

中的信息，来计算分别与其对应的音素部分的基点和持续长度之比。此外，函数提取部 513 将作为该计算结果的比值作为变换函数，将该变换函数和音质 A 的基点及持续长度成组，保存到变换函数数据库 514。

图 26 是表示本实施方式中的函数提取部 513 的一处理例的概略图。

函数提取部 513 从 A 基点数据库 511 和 B 基点数据库 512 中，按分别对应的各音素取得该音素的基点和持续长度。之后，函数提取部 513 对每个音素计算音质 B 的值与音质 A 的值之比。

例如，函数提取部 513 从 A 基点数据库 511 取得音素“m”的持续长度（50ms）、基点 1（2500Hz）、基点 2（4250Hz），并从 B 基点数据库 512 取得音素“m”的持续长度（40ms）、基点 1（2400Hz）、基点 2（4200Hz）。此外，函数提取部 513 将音质 B 的持续长度与音质 A 的持续长度之比（持续长度比）计算为 40/50=0.8，音质 B 的基点 1 与音质 A 的基点 1 之比（基点 1 比）计算为 2400/2500=0.96，音质 B 的基点 2 与音质 A 的基点 2 之比（基点 2 比）计算为 4200/4250=0.988。

当这样计算比值时，函数提取部 513 按每个音素、将音质 A 的持续长度（A 持续长度）、基点 1（A 基点 1）及基点 2（A 基点 2）和计算出的持续长度比、基点 1 比及基点 2 比成组，保存到变换函数数据库 514。

图 27 是表示本实施方式中的函数选择部 515 的一处理例的概略图。

函数选择部 515 按照 A 声音数据 506 所示的各音素，从变换函数数据库 514 检索表示与该音素的基点 1 和基点 2 的组最接近的频率的 A 基点 1 和 A 基点 2 的组。之后，当函数选择部 515 发现该组时，从变换函数数据库 514 中将与该组对应的持续长度比、基点 1 比和基点 2 比作为对该音素的变换函数选择。

例如，当函数选择部 515 从变换函数数据库 514 选择对 A 声音数据 506 所示的音素“m”的变换最佳的变换函数时，从变换函数数据库 514 检索表示与该音素“m”所示的基点 1（2550Hz）及基点 2（4200Hz）最接近的频率的 A 基点 1 及 A 基点 2 的组。即，在变换函数数据库 514 中有对音素“m”的两个变换函数时，函数选择部 515 计算 A 声音数据 506

的音素“m”所示的基点 1 及基点 2（2550Hz，4200Hz）与变换函数数据库 514 的音素“m”所示的 A 基点 1 及 A 基点 2（2500Hz，4250Hz）的距离（类似度）。此外，函数选择部 515 计算 A 声音数据 506 的音素“m”所示的基点 1 及基点 2（2550Hz，4200Hz）与变换函数数据库 514 的音素“m”所示的另一个 A 基点 1 及 A 基点 2（2400Hz，4300Hz）的距离（类似度）。结果，函数选择部 515 将与距离最短的即类似度最高的 A 基点 1 及基点 2(2500Hz,4250Hz)对应的持续长度比(0.8)、基点 1 比(0.96)及基点 2 比（0.988），作为对 A 声音数据 506 的音素“m”的变换函数来选择。

这样，函数选择部 515 对 A 声音数据 506 所示的各音素，选择对该音素最佳的变换函数。即，该函数选择部 515 具备类似度导出机构，对作为单元存储机构的第 1 缓冲器 517 的 A 声音数据 506 中包含的各音素，比较该音素的音响特征（基点 1 和基点 2）、和制作作为函数存储机构的变换函数数据库 514 中所存储的变换函数时使用的音素的音响特征（基点 1 和基点 2），来导出类似度。此外，函数选择部 515 对包含在 A 声音数据 506 中的各音素，选择使用该音素和类似度最高的音素来生成的变换函数。此外，函数选择部 515 生成包含该选择的变换函数、和在变换函数数据库 514 中对应于该变换函数的 A 持续长度、包含 A 基点 1 及 A 基点 2 的变换函数数据 516。

此外，也可以通过按照基点的种类来对距离附加权重，进行优先考虑某个特定种类的基点的位置的接近程度的计算。例如，通过使左右音韵性的低阶共振峰频率的权重较大，能够降低因音质变换而音韵性变形的风险。

图 28 是表示本实施方式中的函数应用部 59 的处理的一例的概略图。

函数应用部 509 通过对 A 声音数据 506 的各音素所表示的持续长度、基点 1 及基点 2，乘上变换函数数据 516 所表示的持续时间长度比、基点 1 比及基点 2 比和由变换率指定部 507 指定的变换率，来校正该 A 声音数据 506 的各音素所示的持续长度、基点 1 及基点 2。此外，函数应用部

509 使 A 声音数据 506 所示的波形数据变形，以与该被校正的持续长度、基点 1 及基点 2 一致。即，本实施方式中的函数应用部 509 对 A 声音数据 506 中包含的各音素，应用由函数选择部 115 选择的变换函数，来改变该音素的音质。

例如，函数应用部 509 在 A 声音数据 506 的音素“u”所表示的持续长度（80ms）、基点 1（3100Hz）及基点 2（4300Hz）上，乘上变换函数数据 516 所表示的持续长度比（1.5）、基点 1 比（0.95）及基点 2 比（1.05）和由变换率指定部 507 指定的变换率 100%。从而， A 声音数据 506 的音素“u”所表示的持续长度（80ms）、基点 1（3000Hz）及基点 2（4300Hz）被修正为持续长度（120ms）、基点 1（2850Hz）及基点 2（4515Hz）。之后，函数应用部 509 对其波形数据进行变形，以使 A 声音数据 506 的波形数据的音素“u”部分的持续长度、基点 1 和基点 2 成为被修正后的持续长度（120ms）、基点 1（2850Hz）及基点 2（4515Hz）。

图 29 是表示本实施方式的音质变换装置的动作的流程图。

首先，音质变换装置取得文本数据 501（步骤 S500）。音质变换装置对该取得的文本数据 501 进行语言分析或词素分析等，并根据该分析结果生成韵律（步骤 S502）。

当生成韵律时，音质变换装置通过根据该韵律从 A 单元数据库 510 选择并连接音素，来生成表示音质 A 的声音的 A 声音数据 506（步骤 S504）。

音质变换装置确定 A 声音数据中包含的最初音素的基点（步骤 S506），将基于与该基点最近的基点生成的变换函数作为对该音素最佳的变换函数，从变换函数数据库 514 中选择（步骤 S508）。

在此，音质变换装置判断是否对在步骤 S504 生成的 A 声音数据中包含的所有音素都选择了变换函数（步骤 S510）。在判断为没有被选择时（步骤 S510 的“否”），音质变换装置对 A 声音数据 506 中包含的下一个音素重复执行步骤 S506 后的处理。另一方面，在判断为被选择时（步骤 S510 的“是”），音质变换装置通过将所选择的变换函数适用于 A 声音数据 506，

将该A声音数据506变换为音质B的声音所示的已变换声音数据508(步骤S512)。

在这样的本实施方式中，通过对A声音数据506的音素使用根据与该音素的基点最近的基点来生成的变换函数，将A声音数据506所表示的声音的音质从音质A变换为音质B。因此，在本实施方式中，例如A声音数据506中有多个相同的音素、并且这些音素的音响特征不同时，不会像现有例那样不管音响特征不同将相同的变换函数用于这些音素，而应用对应于该音响特征的变换函数，能够适当地变换A声音数据506所示的声音的音质。

此外，在本实施方式中，用称作基点的代表值简单地表示了音响特征，因此，在从变换函数数据库514选择变换函数时，不进行复杂的运算处理即可简单且迅速并适当地选择变换函数。

此外，在以上的方法中，将各音素内的各基点的位置或对各音素内的各基点位置的倍率设定为恒定值，但是也可以分别光滑地内插到音素之间。例如，图28中，音素“u”的中心位置中的基点1的位置是3000Hz、音素“m”的中心位置中为2550Hz，但是在其中间时刻，考虑到基点1的位置为（3000+2550）/2=0.955，也可以进行变形，以使声音在该时刻的短时间频谱的2775Hz附近对准进入到2775×0.955=2650.125Hz附近。

再有，在上述方法中，通过使声音的频谱形状变形来进行了音质变换，但也可以通过变换模型基本（モデルベース）声音合成法的模型参数值来进行音质变换。该情况下，可以不把基点位置提供到声音频谱上，而代之把各波形参数提供到各模型参数的时间系列变化图表上。

此外，在上述方法中，以对全部音素使用共同种类的基点为其前提，但是也可以改变根据音素的种类使用的基点的种类。例如，在元音中，以共振峰频率为基础定义基点信息的情况较有效，但是在无声辅音中，由于共振峰定义自身的物理意义较少，因此，也可以考虑与适用于元音的共振峰分析分开而独立地提取频谱上的特征点（峰值等），并设定为基点信息，这种情况也是有效的。此时，在元音部和无声辅音部设定的基

点信息的个数（维数）相互不同。

（变形例 1）

在上述实施方式的方式中，以音质变换为音素单位进行，但也能够以比单词单位和重音语句单位等更长的单位来进行。尤其是决定韵律的基本频率和持续长度的信息很难仅用音素单位来完成处理，因此，用变换目标的音质决定对文本整体的韵律信息，并通过进行与变换源音质中的韵律信息的替换或渐变（morphing）来进行变形。

即，本变形例中的音质变换装置通过分析文本数据 501，来生成与将音质 A 靠近音质 B 的中间音质对应的韵律信息（中间韵律信息），并从 A 单元数据库 510 选择与该中间韵律信息对应的音素，来生成声音数据 506。

图 30 是表示本变形例的音质变换装置结构的结构图。

本变形例的音质变换装置具备生成与从音质 A 靠近音质 B 的音质对应的中间韵律信息的韵律生成部 503a。

该韵律生成部 503a 具备：A 韵律生成部 601、B 韵律生成部 602、中间韵律生成部 603。

A 韵律生成部 601 生成包含附加在音质 A 的声音上的重音或各音素的持续长度等的 A 韵律信息。

B 韵律生成部 602 生成包含附加在音质 B 的声音上的重音或各音素的持续长度等的 B 韵律信息。

中间韵律生成部 603 根据分别由 A 韵律生成部 601 及 B 韵律生成部 602 生成的 A 韵律信息及 B 韵律信息、和由变换率指定部 507 指定的变换率进行计算，来生成与将音质 A 靠近音质 B 该变换率程度的音质对应的中间韵律信息。再有，变换率指定部 507 对中间韵律生成部 603 指定与对函数应用部 509 指定的变换率相同的变换率。

具体来说，中间韵律生成部 603 按照由变换率指定部 507 指定的变形率，对分别与 A 韵律信息和 B 韵律信息对应的音素计算持续长度的中间值和各时刻中的基本频率的中间值，并生成表示这些计算结果的中间韵律信息。之后，中间韵律生成部 603 将该生成的中间韵律信息输出到

单元选择部 505。

通过以上的结构，能够进行将可在音素单位内变形的共振峰频率等的变形和文本单位内的变形有效的韵律信息变形组合的音质变换处理。

此外，在本变形例中，根据中间韵律信息选择音素，并生成了 A 声音数据 506，因此，在函数应用部 509 将 A 声音数据 506 变换为已变换声音数据 508 时，可防止无理的音质变换引起的音质的恶化。

（变形例 2）

在上述方法中，通过在各音素的中心位置定义基点，来稳定地表现各音素的音响特征，但是也可以将基点定义为音素内的各共振峰频率的平均值、音素内的各频带的频谱强度的平均值、这些值的分散值等。即，也可以通过按照在声音识别技术中一般使用的 HMM 音响模型的形式定义基点，极端单元侧模型的各状态变量和变换函数侧模型的各状态变量之间的距离，来选择最佳的函数。

与上述实施方式比较，该方法中由于基点信息包含更多的信息，所以具有能够选择更适合的函数的优点，但是有如下缺点：为了基点信息的大小变大而使得选择处理的负荷变大，保持基点信息的各数据库的大小也变大。再有，在从 HMM 音响模型生成声音的 HMM 声音合成装置中，具有能够将单元数据和基点信息共同化的优良效果。即，只要比较表示各变换函数的生成源声音的特征的 HMM 的各状态变量和所使用的 HMM 音响模型的各状态变量，来选择最佳的变换函数即可。表示各变量的生成源声音的特征的 HMM 的各状态变量在用于合成的 HMM 音响中识别生成源声音，只要在各音素内的对应于各 HMM 状态的部分计算音响特征量的平均或分散值就可以。

（变形例 3）

本实施方式是将文本数据 51 作为输入来接受并输出声音的声音合成装置中组合音质变换功能的方式，但也可以将声音作为输入来接受、并利用输入声音的自动标注来生成标识信息、在各音素中心提取频谱峰值点来自动生成基点信息。这样，能够将本发明的技术作为声音转换装置

来使用。

图 31 是表示本变形例的音质变换装置的结构的结构图。

本变形例的音质变换装置包括：上述实施方式的图 23 所示的文本分析部 502、韵律生成部 503、单元连接部 504、单元选择部 505，以及代替 A 单元数据库 510 的 A 声音数据生成部 700。该 A 声音数据生成部 700 把音质 A 的声音作为输入声音来取得，并生成与该输入声音对应的 A 声音数据 506。即，本变形例中，A 声音数据生成部 700 构成为生成 A 生意数据 506 的生成机构。

A 声音数据生成部 700 包括麦克风 705、标注部 702、音响特征分析部 703、标注用音响模型 704。

麦克风 705 收集输入声音，并生成表示该输入声音的波形的 A 输入声音波形数据 701。

标注部 702 参照标注用音响模型 704，对 A 输入声音波形数据 701 进行音素的标注。从而生成对该 A 输入声音波形数据 701 种包含的音素的标签信息。

音响特征分析部 703 通过提取由标注部 702 标注的各音素中心点(时间轴中心）中的频谱峰值点（共振峰频率），来生成基点信息。此外，音响特征分析部 703 生成包括所生成的基点信息、标注部 702 生成的标签信息和 A 输入声音波形数据 701 的 A 声音数据 506，并存储到第 1 缓冲器 517。

从而，在本变形例中，能够变换所输入的声音音质。

此外，用实施方式和其变形例来对本发明进行说明，但是并不限定于此。

例如，在本实施方式及其变形例中，如基点 1 和基点 2，将基点数设定为两个，并如基点 1 比和基点 2 比那样，将变换函数中的基点比的个数设定为两个，但是也可以将基点和基点比的个数分别设定为 1 个，也可以设定为 3 个以上。通过增加基点和基点比的个数，能够对音素选择更加合适的变换函数。

产业上的可利用性

本发明的声音合成装置具有可适当地变换音质的效果，并且，可用于例如汽车导航系统、家庭用电器产品等娱乐性较高的声音接口、分开使用各种音质的同时进行合成音的信息提供的装置、以及应用程序等中，尤其是在需要声音的感情表现的邮件文章的读取或要求表现讲话者的性别的代理应用程序等用途中有用。此外，通过组合声音的自动标注技术，也可以应用到可按所希望的歌手的音质来唱歌的卡拉 OK 装置、或以个人秘密保护等为目的的声音转换等中。

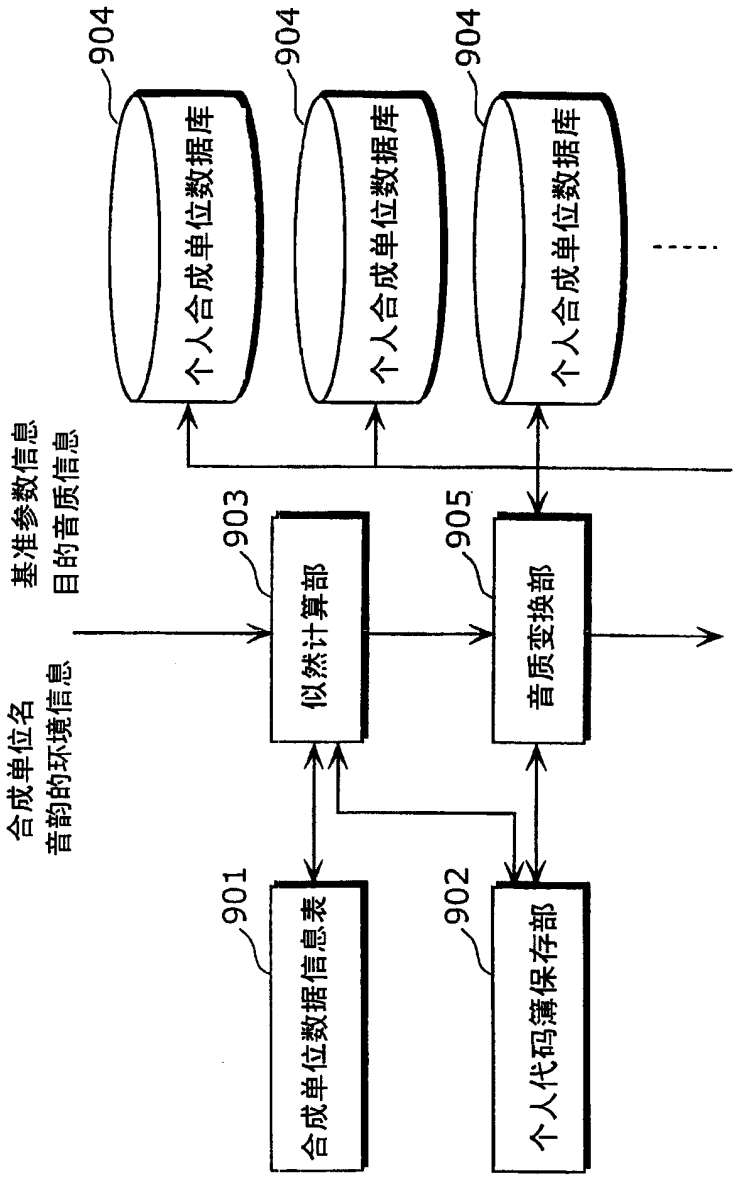

904
个人合成单位数据库
904
个人合成单位数据库
904
个人合成单位数据库
基准参数信息
目的音质信息
903
似然计算部
905
音质变换部
合成单位名
音韵的环境信息
901
合成单位数据信息表
902
个人代码簿保存部

图1

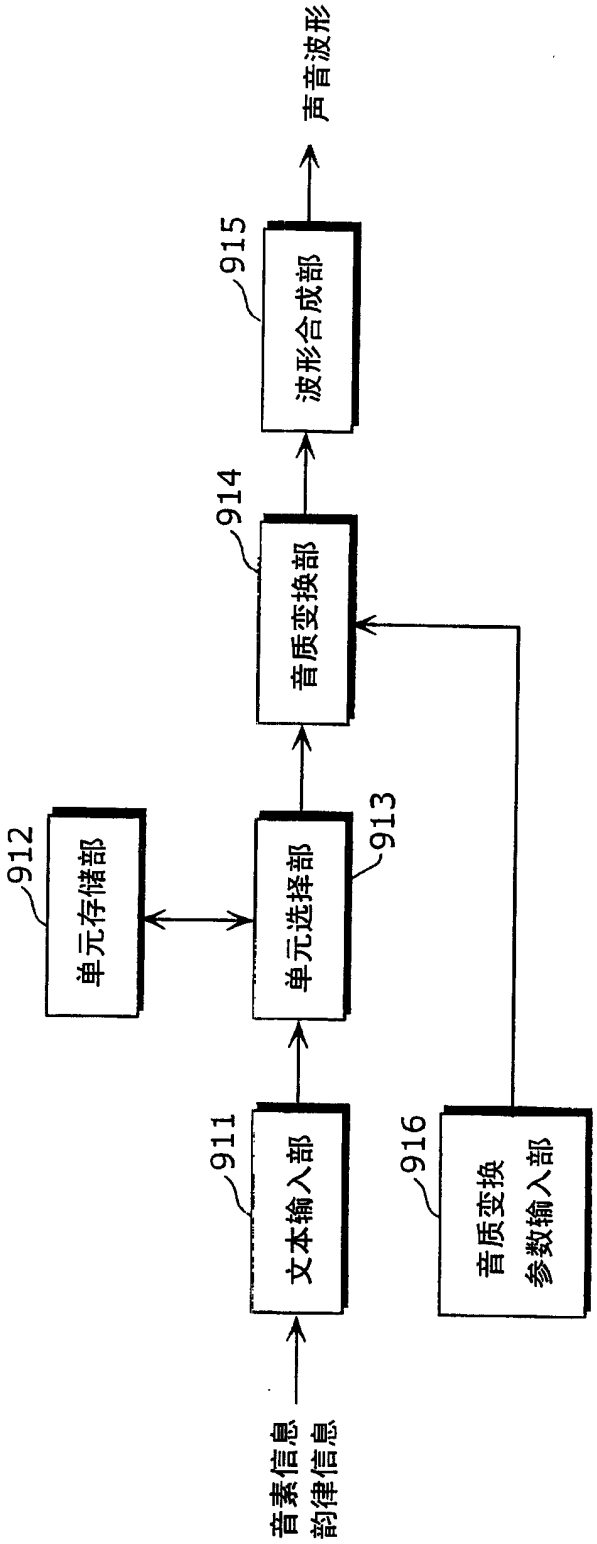

音素信息
韵律信息
911
文本输入部
912
单元存储部
913
单元选择部
914
音质变换部
915
波形合成部
声音波形
916
音质变换
参数输入部

图2

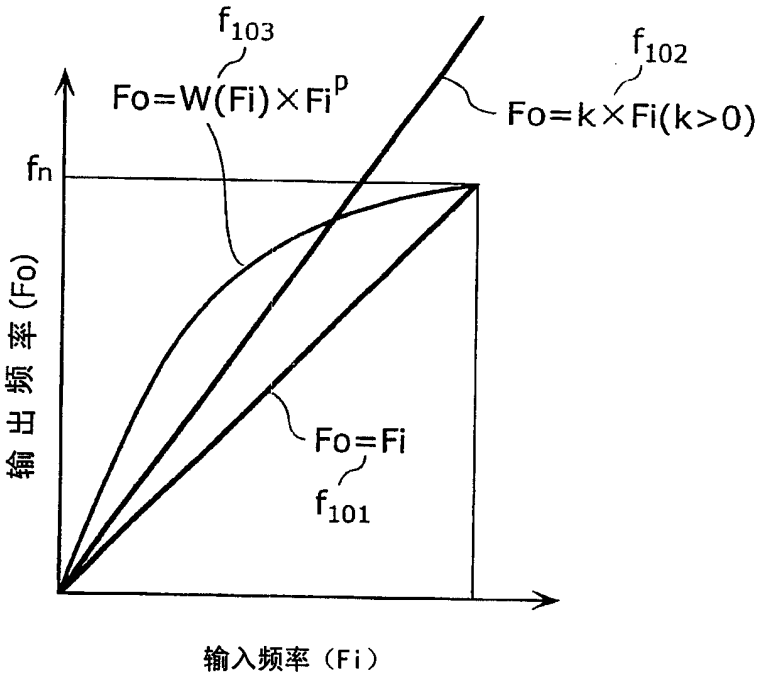


图3

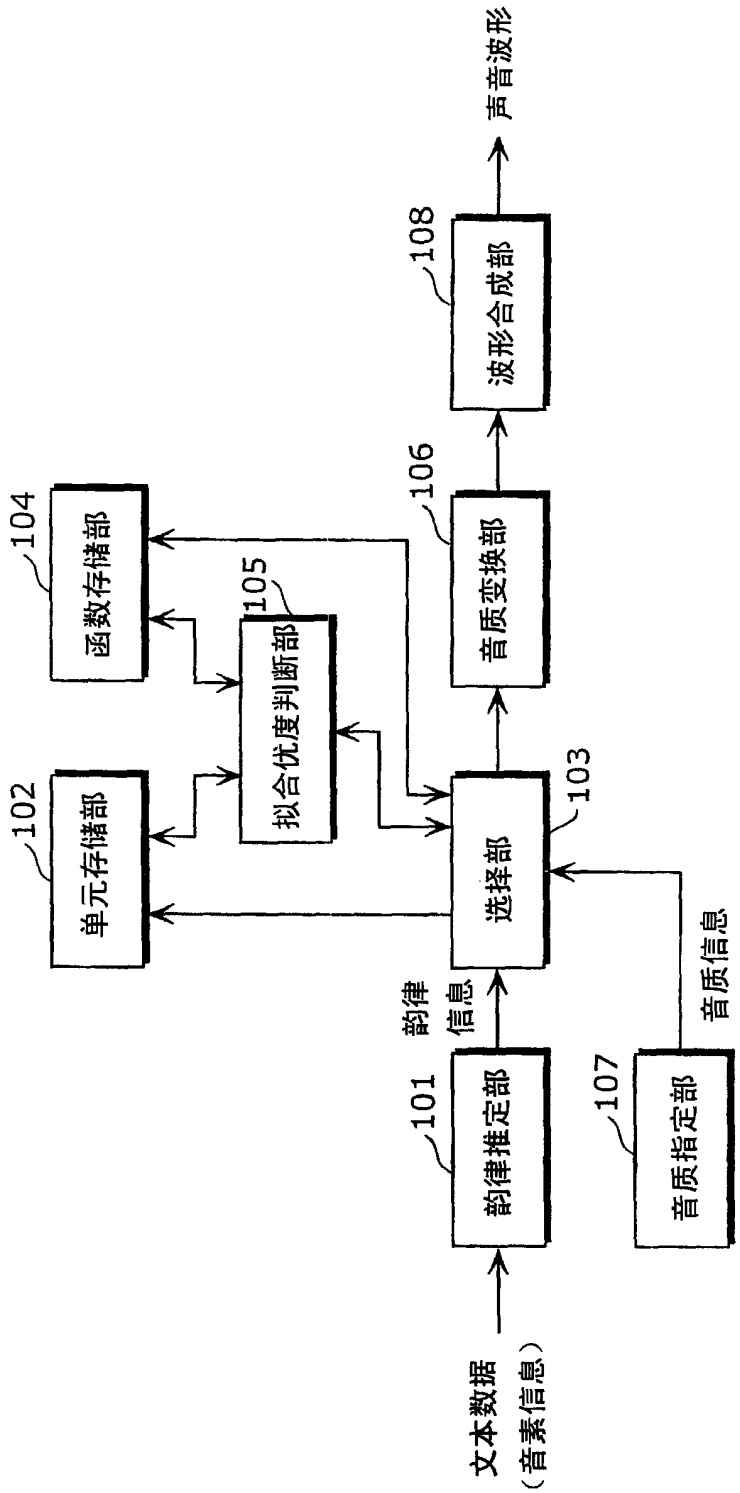

文本数据
（音素信息）
韵律推定部
101
韵律信息
选择部
103
单元存储部
102
函数存储部
104
拟合优度判断部
105
音质变换部
106
波形合成部
108
声音波形
音质指定部
107
音质信息

图4

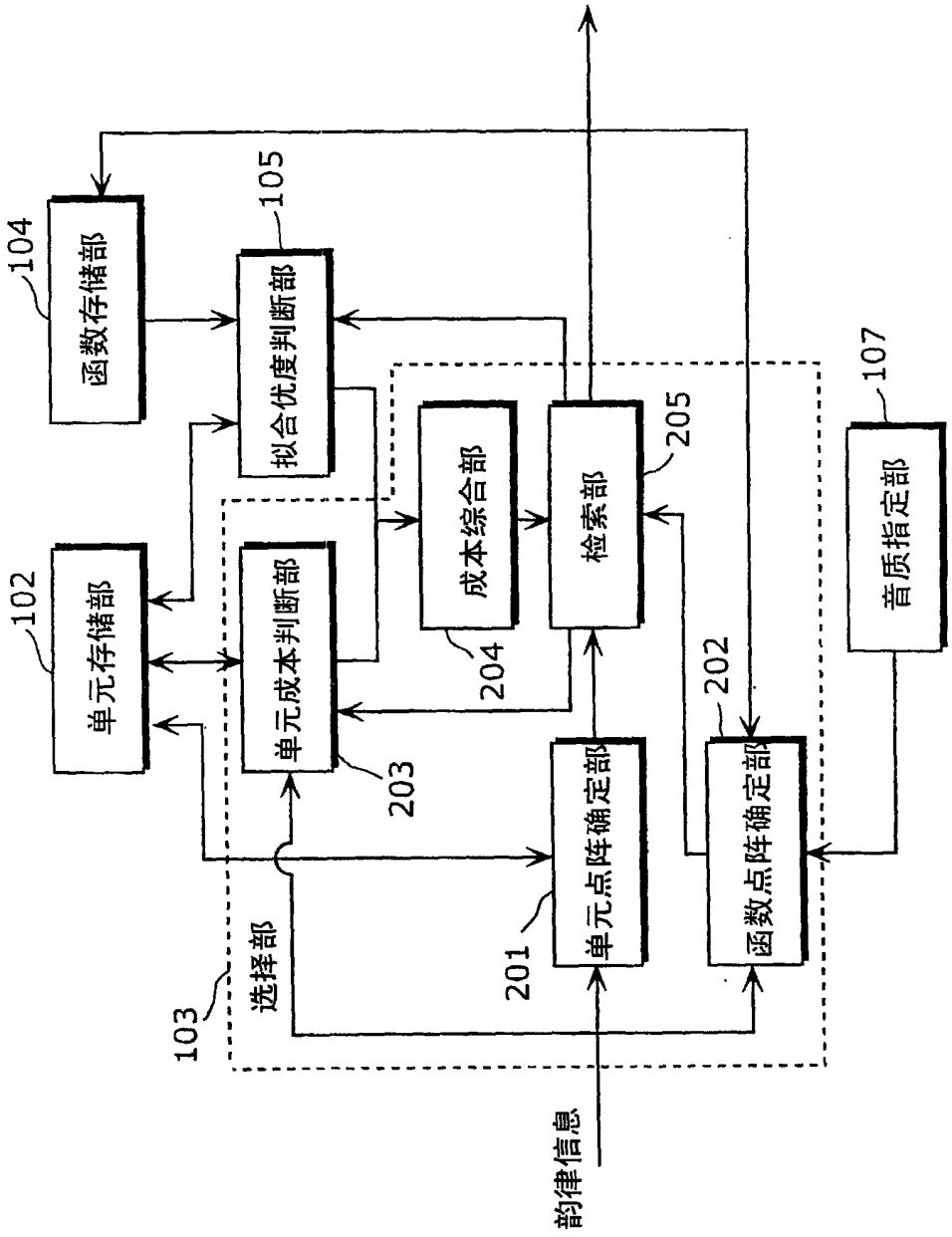

韵律信息
选择部
103
单元点阵确定部
201
函数点阵确定部
202
单元成本判断部
203
成本综合部
204
检索部
205
单元存储部
102
拟合优度判断部
105
函数存储部
104
音质指定部
107

图5

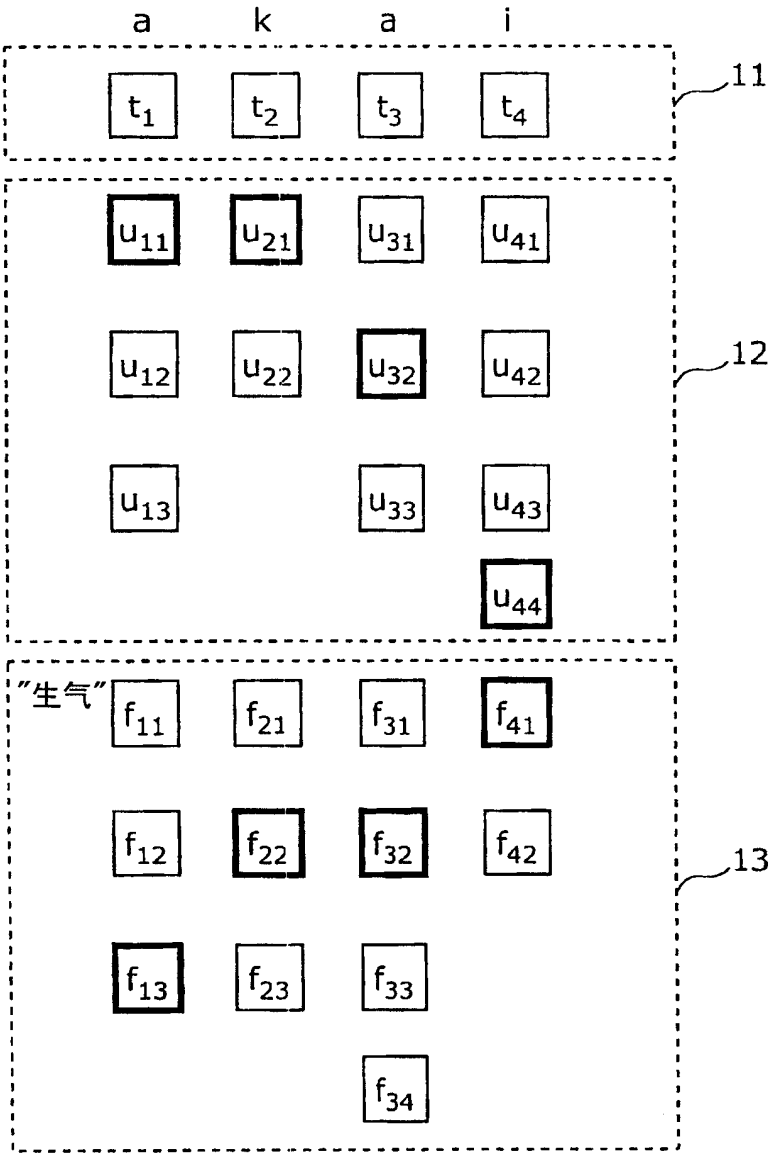

a k a i
11
12
13
"生气"
t1 t2 t3 t4
u11 u21 u31 u41
u12 u22 u32 u42
u13 u33 u43
u44
f11 f21 f31 f41
f12 f22 f32 f42
f13 f23 f33
f34

图6

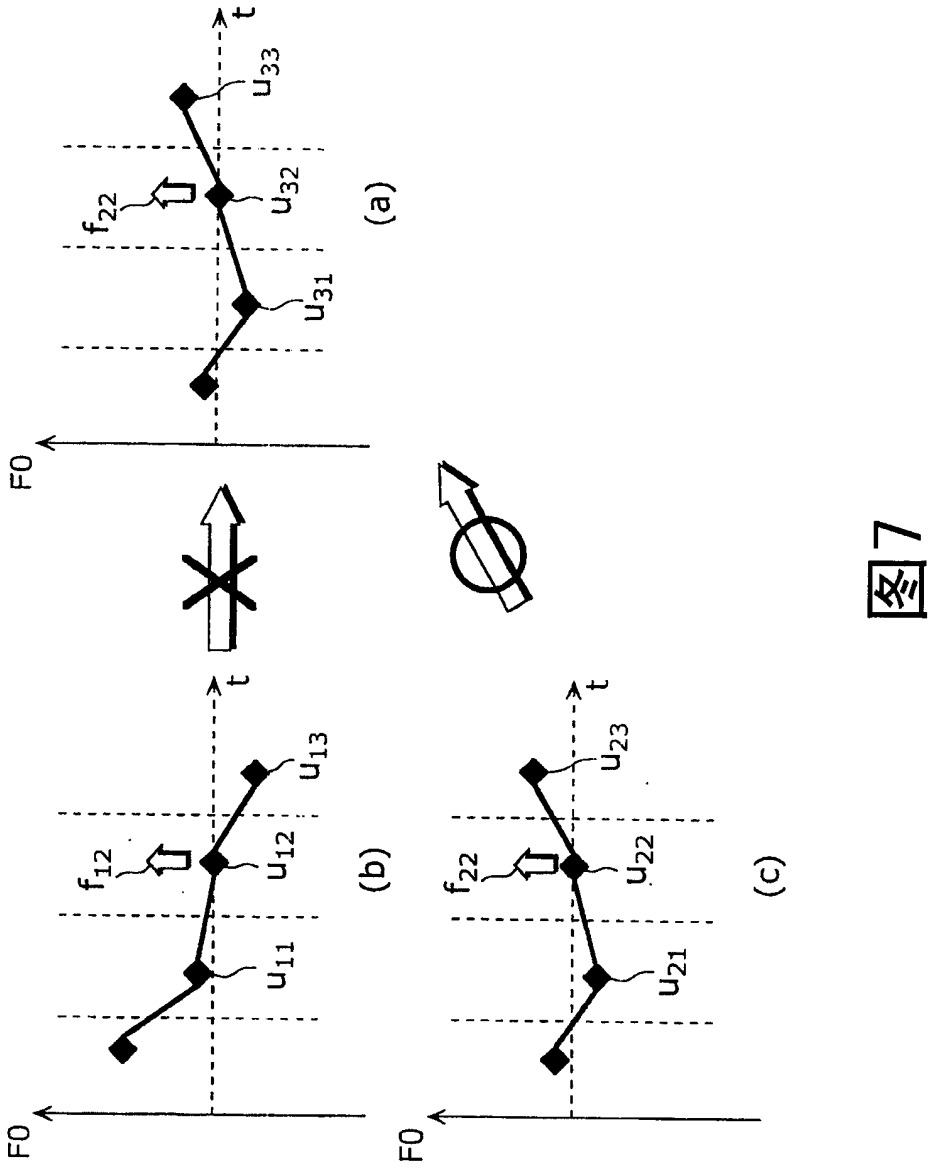


图7

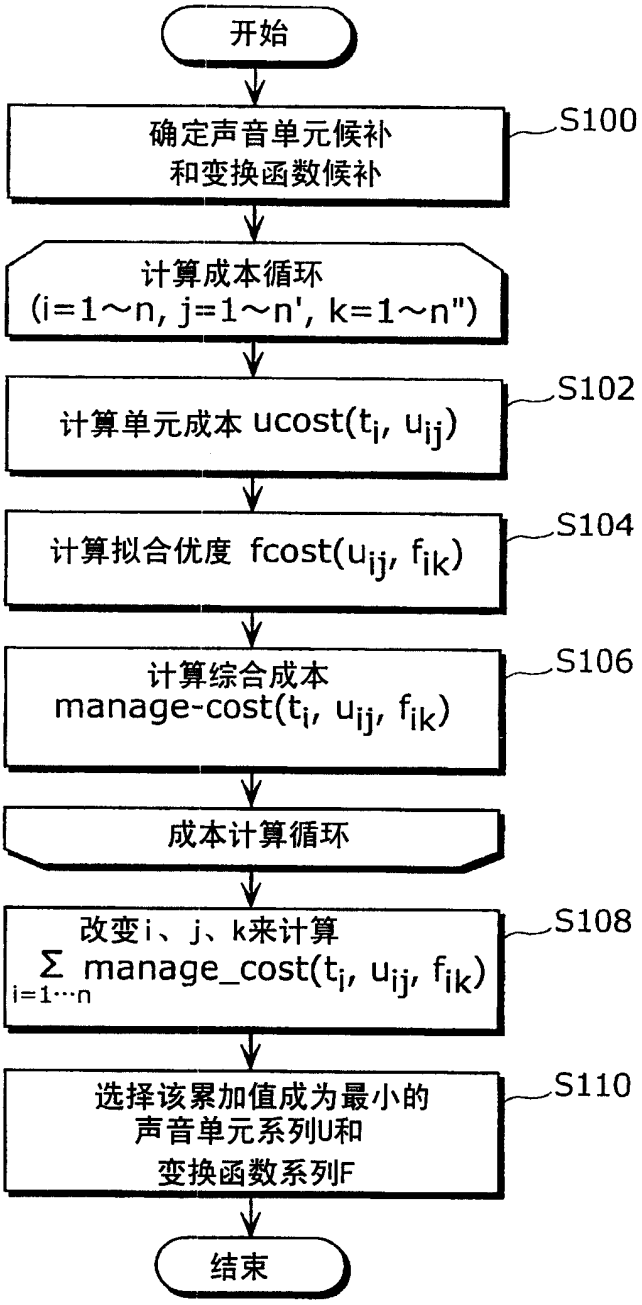

开始
确定声音单元候补
和变换函数候补
S100
计算成本循环
(i=1～n, j=1～n', k=1～n")
计算单元成本 ucost(ti, uij)
S102
计算拟合优度 fcost(uij, fik)
S104
计算综合成本
manage-cost(ti, uij, fik)
S106
成本计算循环
改变i、j、k来计算
Σ manage_cost(ti, uij, fik)
i=1…n
S108
选择该累加值成为最小的
声音单元系列U和
变换函数系列F
S110
结束

图8

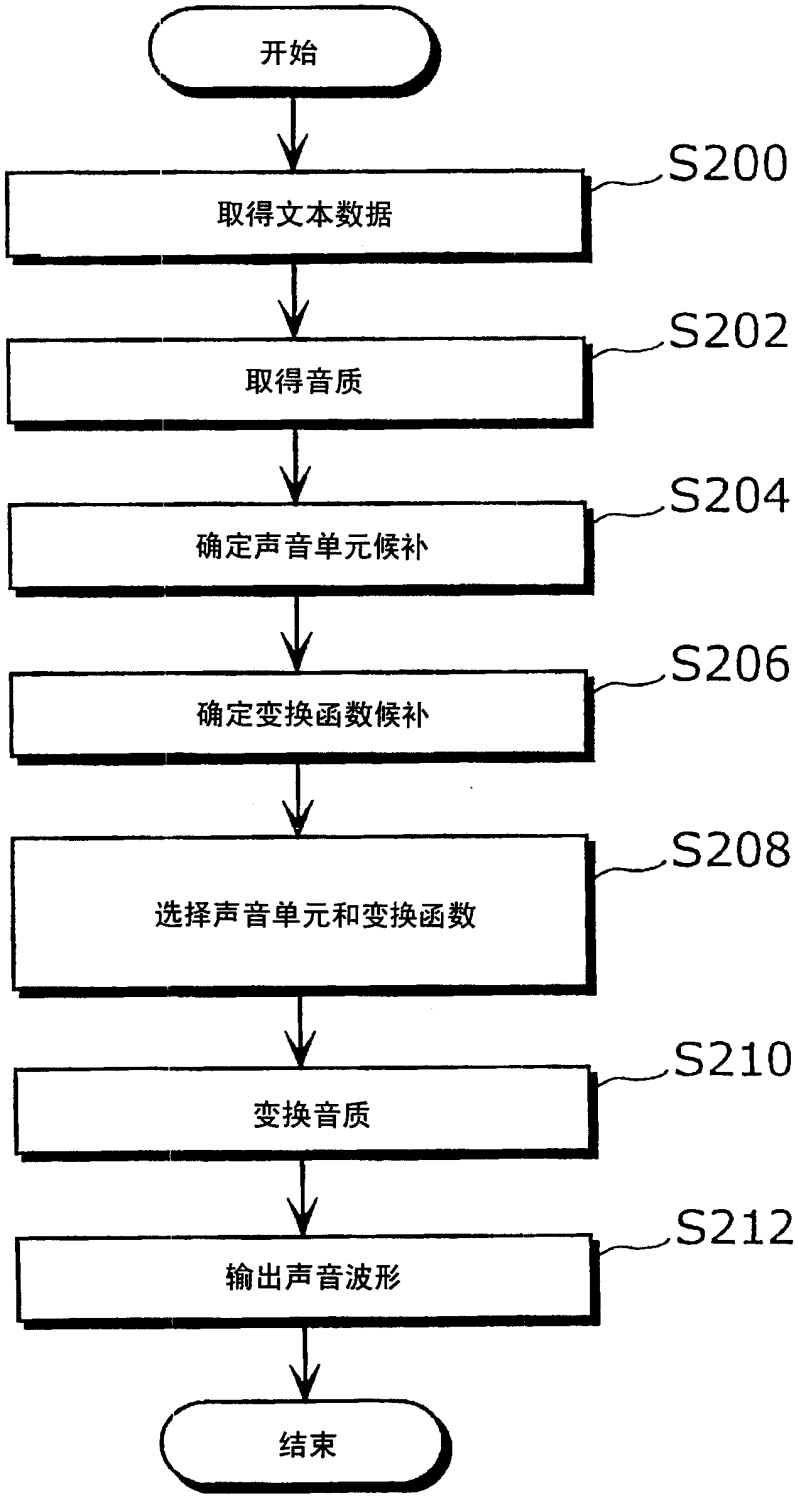

开始
S200
取得文本数据
S202
取得音质
S204
确定声音单元候补
S206
确定变换函数候补
S208
选择声音单元和变换函数
S210
变换音质
S212
输出声音波形
结束

图9

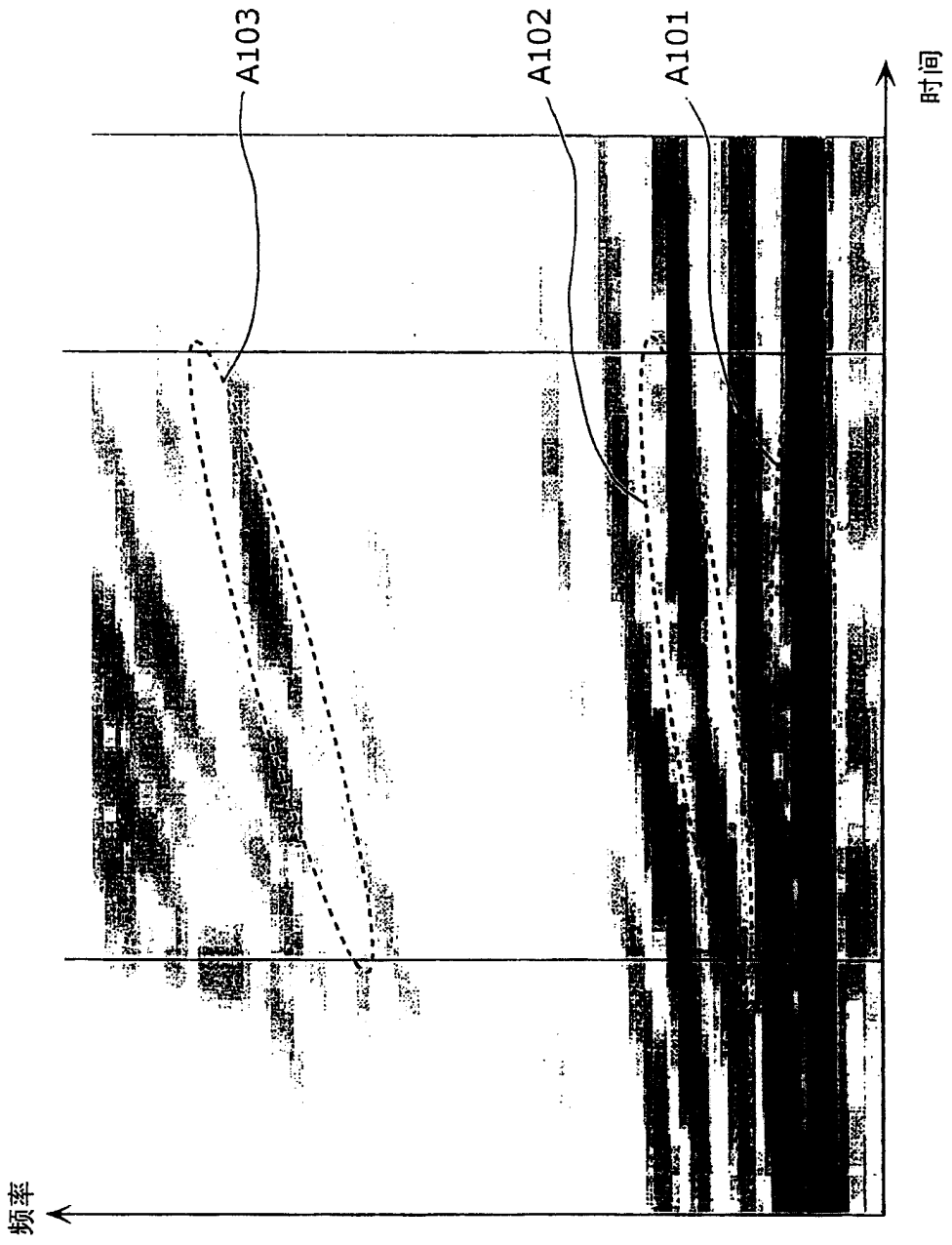


图10

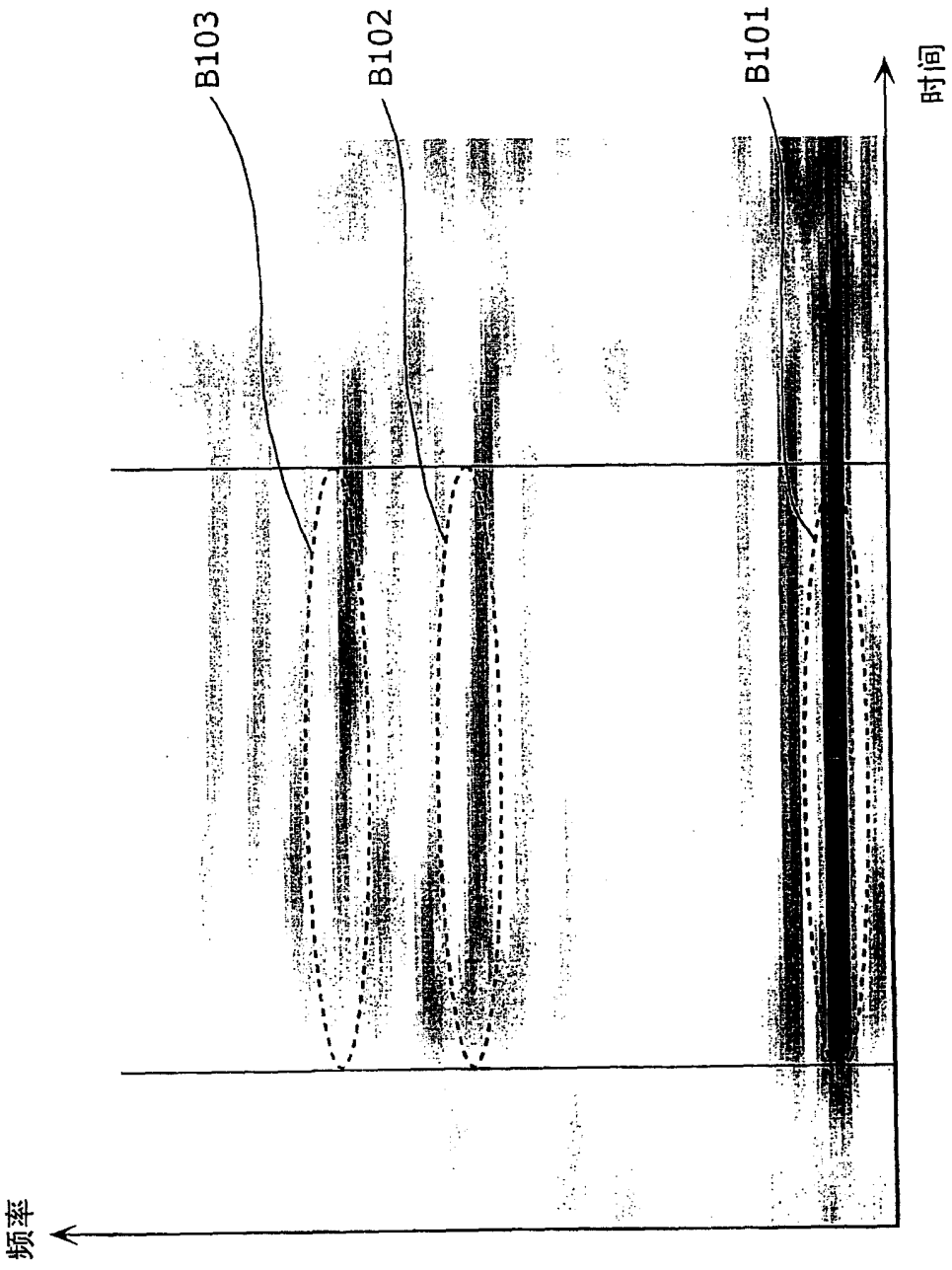


图11

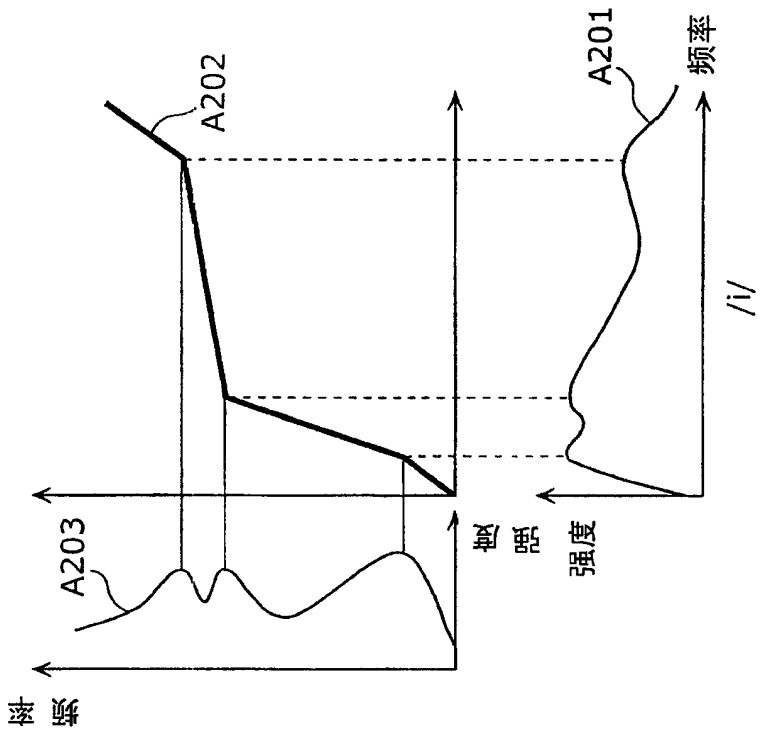


图12A

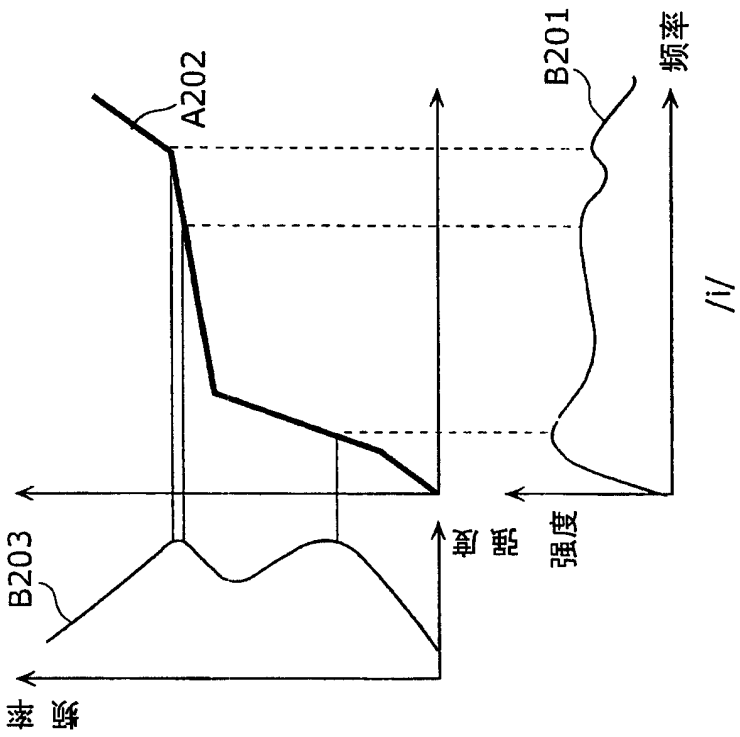


图12B

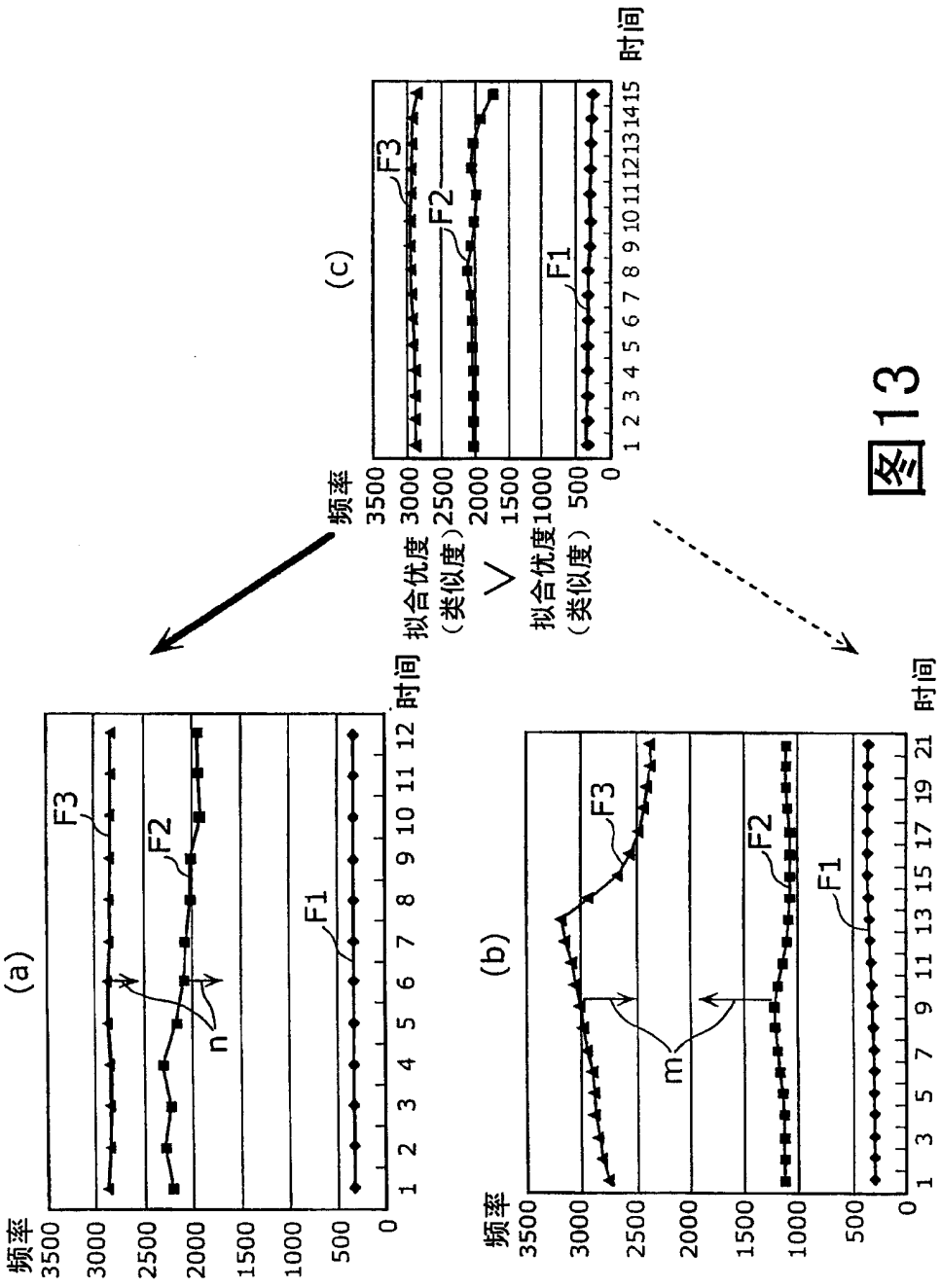


图13

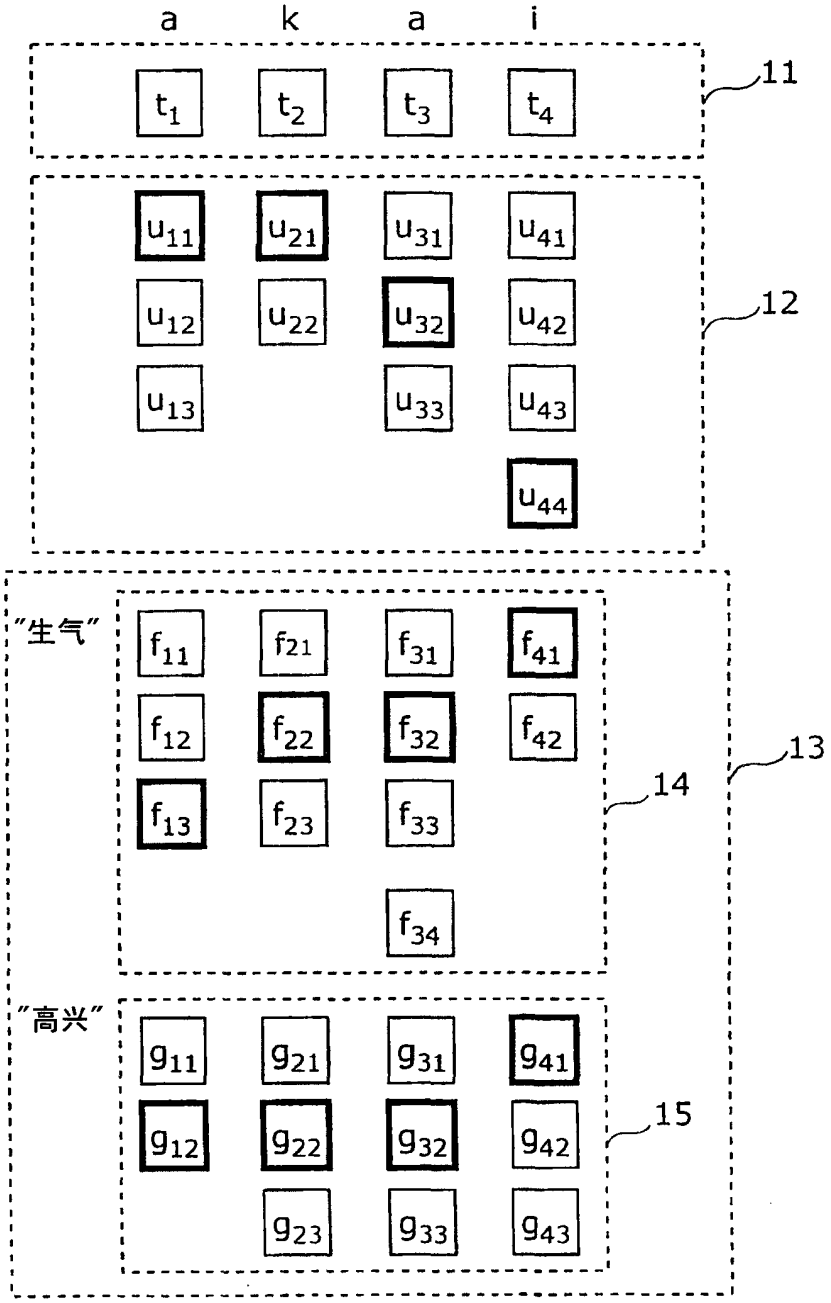


图14

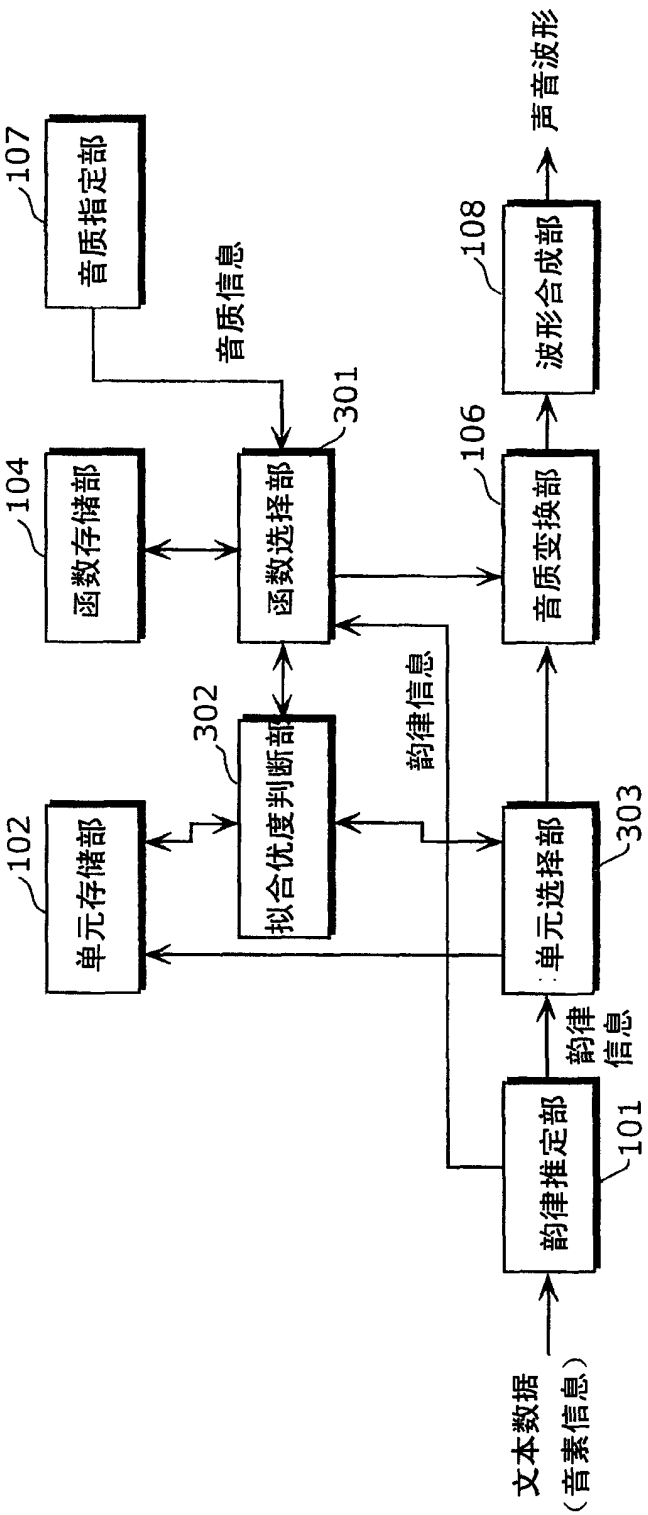

文本数据
（音素信息）
韵律推定部
101
韵律信息
单元选择部
303
拟合优度判断部
302
韵律信息
单元存储部
102
函数选择部
301
函数存储部
104
音质信息
音质指定部
107
音质变换部
106
波形合成部
108
声音波形

图15

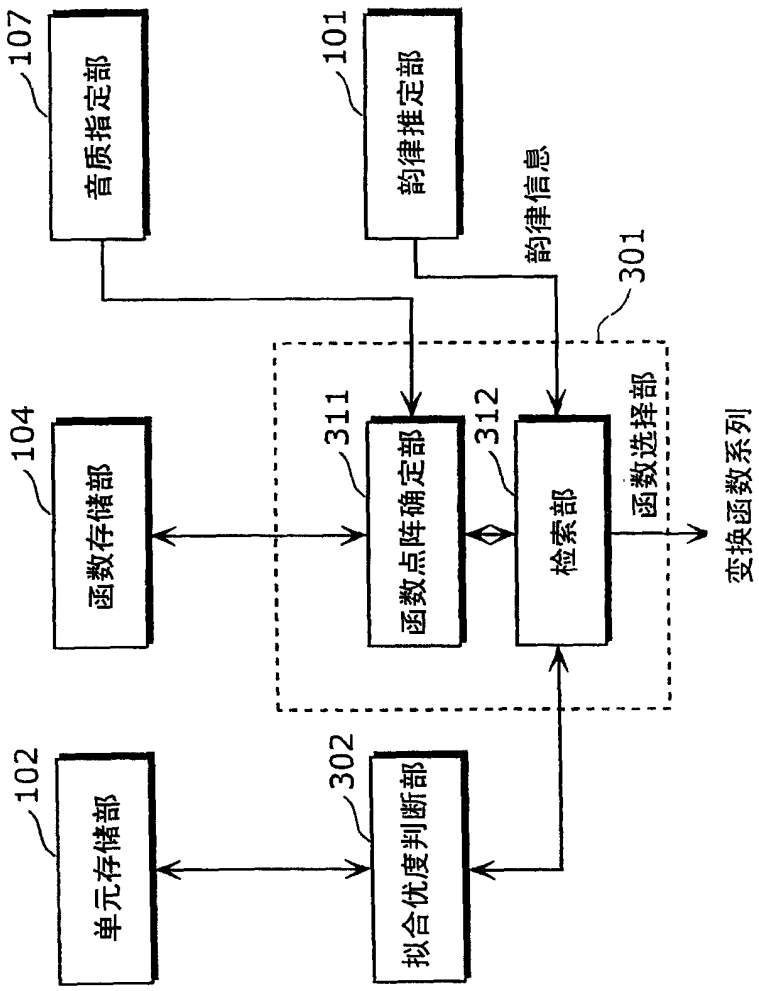

107
音质指定部
101
韵律推定部
韵律信息
301
104
函数存储部
311
函数点阵确定部
312
检索部
函数选择部
变换函数系列
102
单元存储部
302
拟合优度判断部

图16

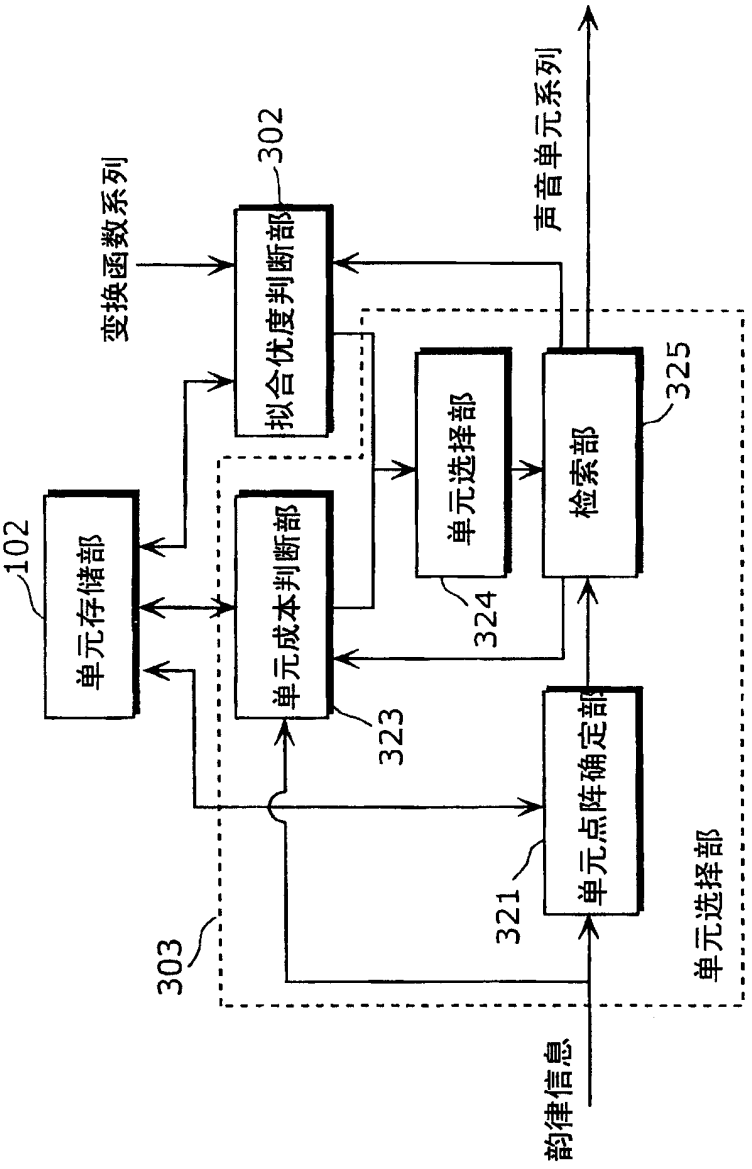

变换函数系列
302
拟合优度判断部
声音单元系列
102
单元存储部
单元成本判断部
单元选择部
检索部
325
323
324
303
321
单元点阵确定部
单元选择部
韵律信息

图17

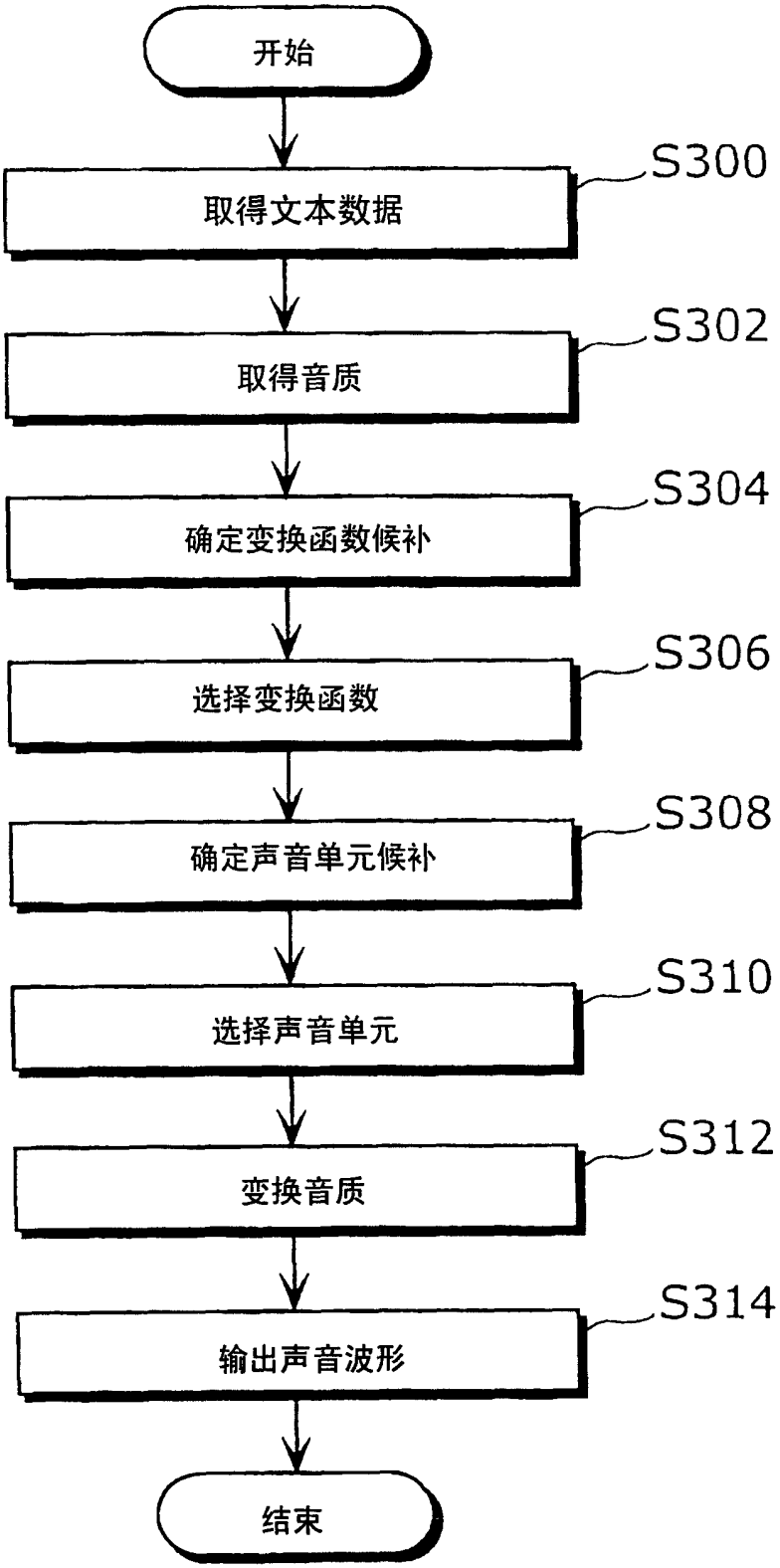

开始
S300
取得文本数据
S302
取得音质
S304
确定变换函数候补
S306
选择变换函数
S308
确定声音单元候补
S310
选择声音单元
S312
变换音质
S314
输出声音波形
结束

图18

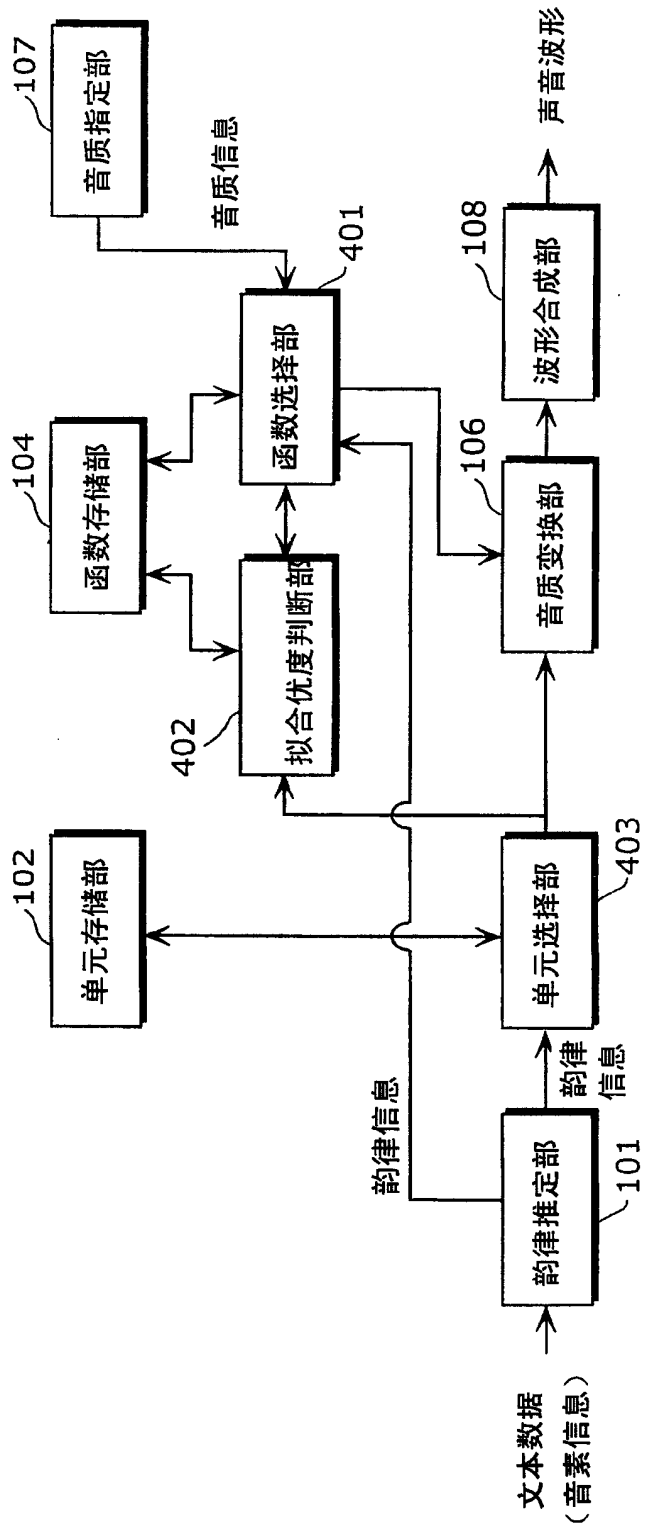

文本数据
（音素信息）
韵律推定部
101
韵律信息
韵律信息
单元选择部
403
单元存储部
102
拟合优度判断部
402
函数选择部
401
函数存储部
104
音质指定部
107
音质信息
音质变换部
106
波形合成部
108
声音波形

图19

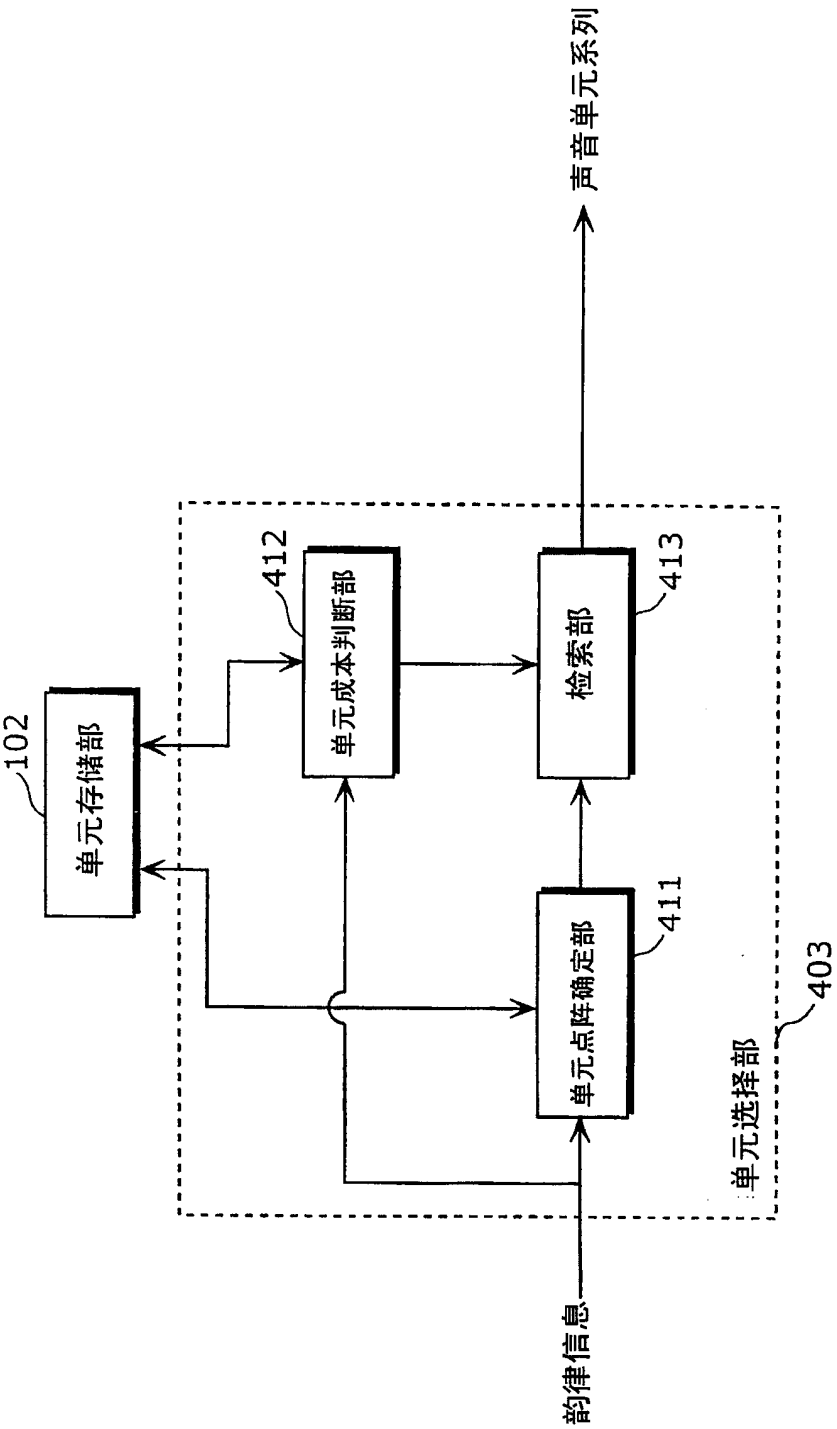

韵律信息
单元选择部
403
单元点阵确定部
411
检索部
413
单元成本判断部
412
单元存储部
102
声音单元系列

图20

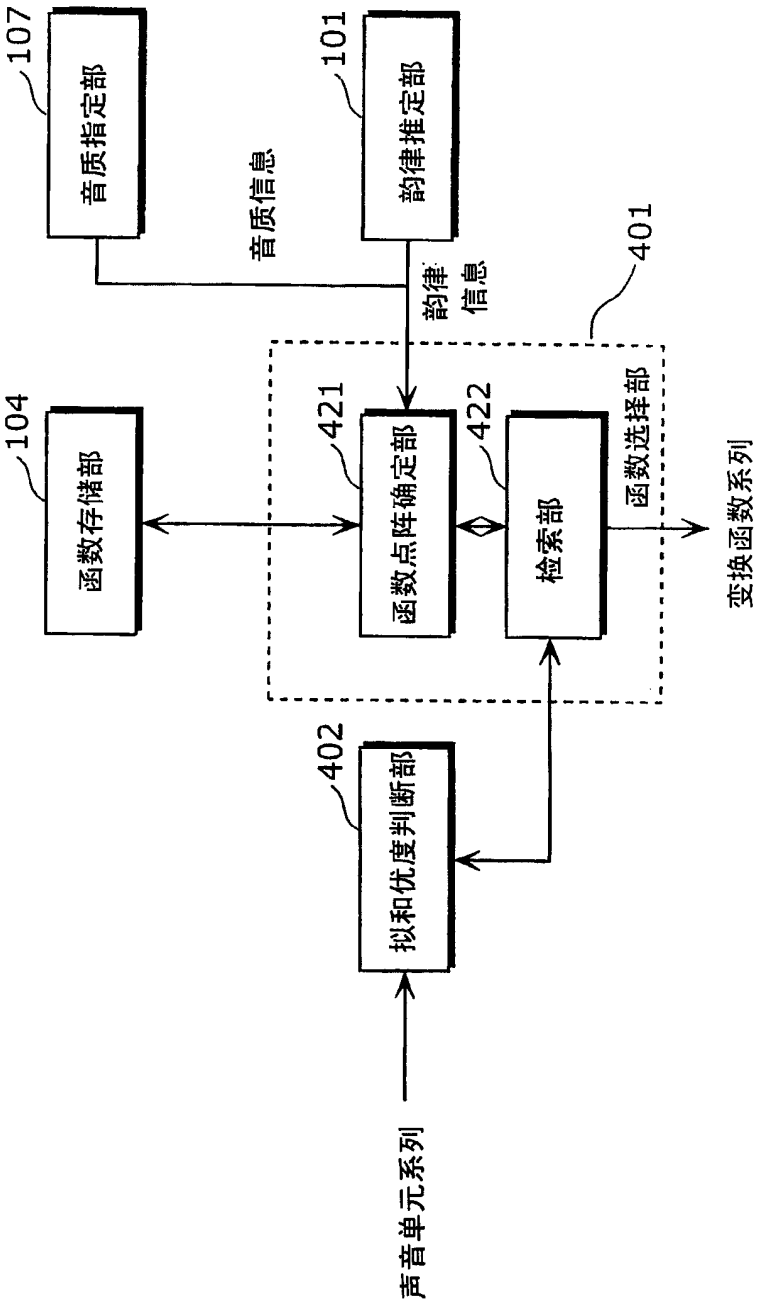

107
音质指定部
101
韵律推定部
音质信息
韵律信息
401
104
函数存储部
421
函数点阵确定部
422
检索部
函数选择部
变换函数系列
402
拟和优度判断部
声音单元系列

图21

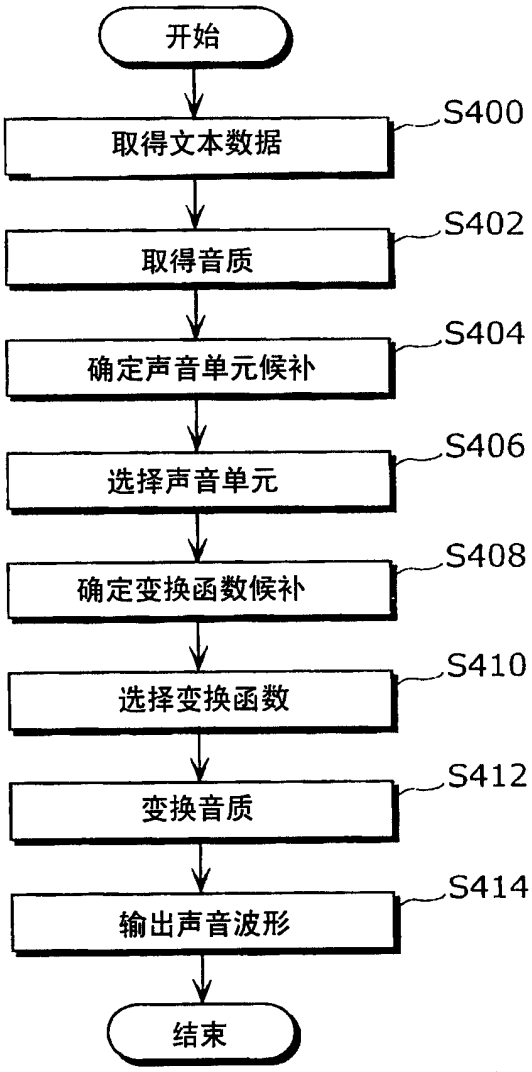

开始
取得文本数据
S400
取得音质
S402
确定声音单元候补
S404
选择声音单元
S406
确定变换函数候补
S408
选择变换函数
S410
变换音质
S412
输出声音波形
S414
结束

图22

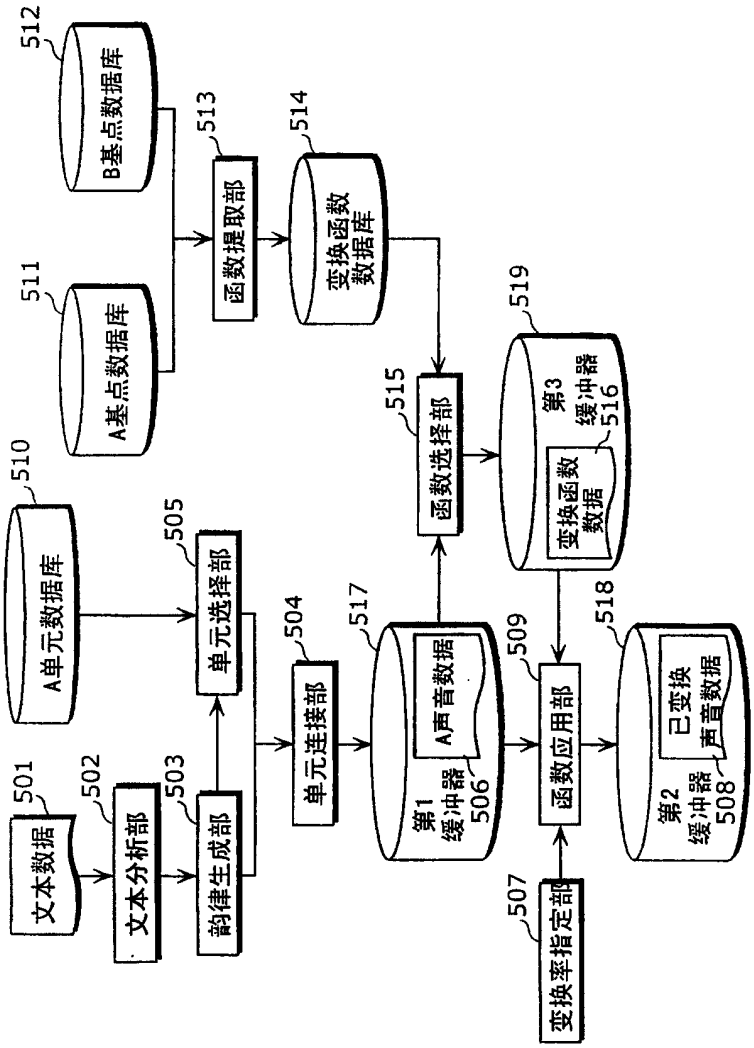

512
B基点数据库
513
函数提取部
514
变换函数数据库
511
A基点数据库
519
第3缓冲器
516
变换函数数据
515
函数选择部
510
A单元数据库
505
单元选择部
504
517
A声音数据
第1缓冲器
506
509
函数应用部
518
已变换声音数据
第2缓冲器
508
501
文本数据
502
文本分析部
503
韵律生成部
507
变换率指定部

图23

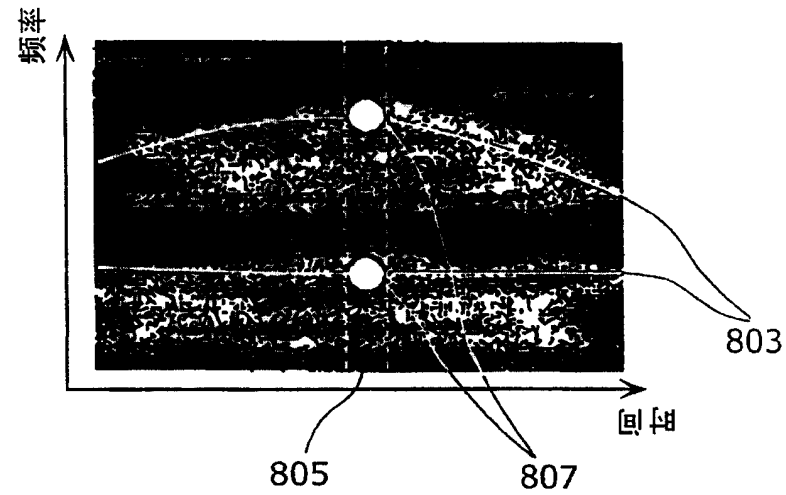

频率
时间
803
805
807

图24A

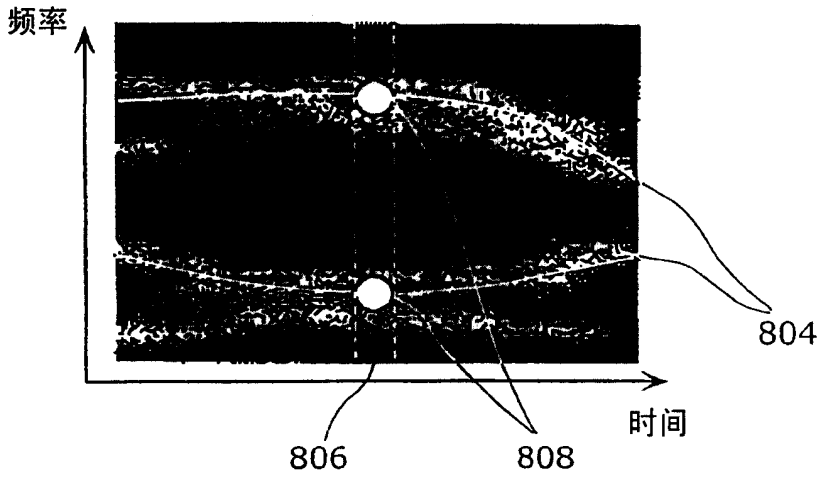

频率
804
时间
806
808

图24B

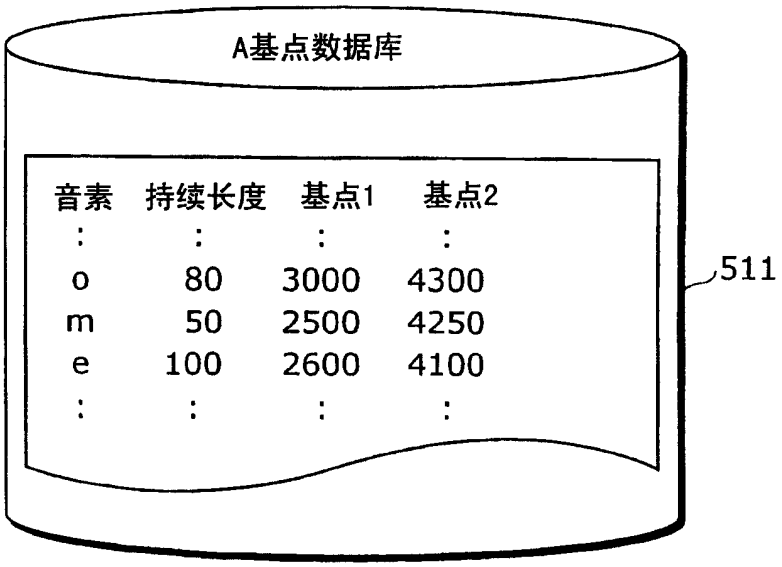

A基点数据库
音素 持续长度 基点1 基点2
: : : :
o 80 3000 4300
m 50 2500 4250
e 100 2600 4100
: : : :
511

图25A

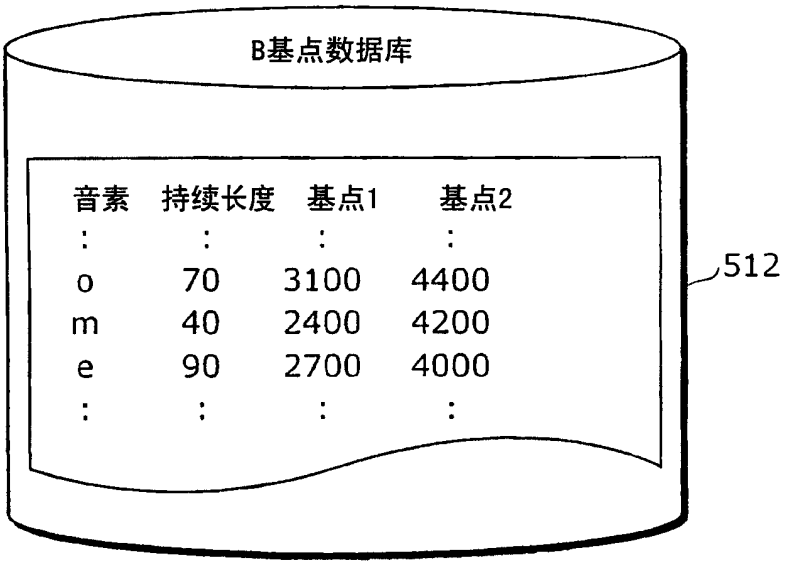

B基点数据库
音素 持续长度 基点1 基点2
: : : :
o 70 3100 4400
m 40 2400 4200
e 90 2700 4000
: : : :
512

图25B

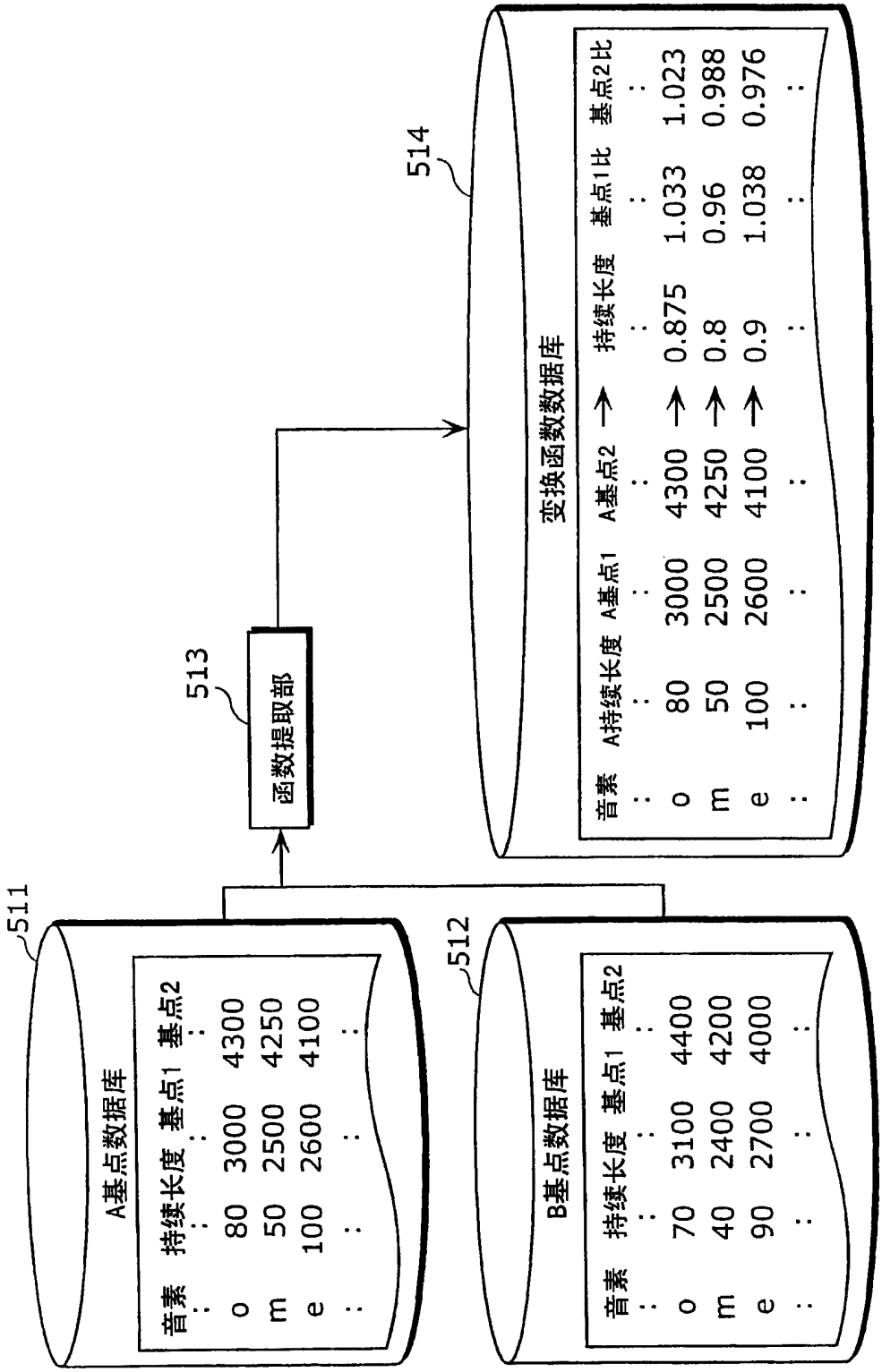

511
A基点数据库
音素 持续长度 基点1 基点2
o 80 3000 4300
m 50 2500 4250
e 100 2600 4100
512
B基点数据库
音素 持续长度 基点1 基点2
o 70 3100 4400
m 40 2400 4200
e 90 2700 4000
513
函数提取部
514
变换函数数据库
音素 A持续长度 A基点1 A基点2 → 持续长度 基点1比 基点2比
o 80 3000 4300 → 0.875 1.033 1.023
m 50 2500 4250 → 0.8 0.96 0.988
e 100 2600 4100 → 0.9 1.038 0.976

图26

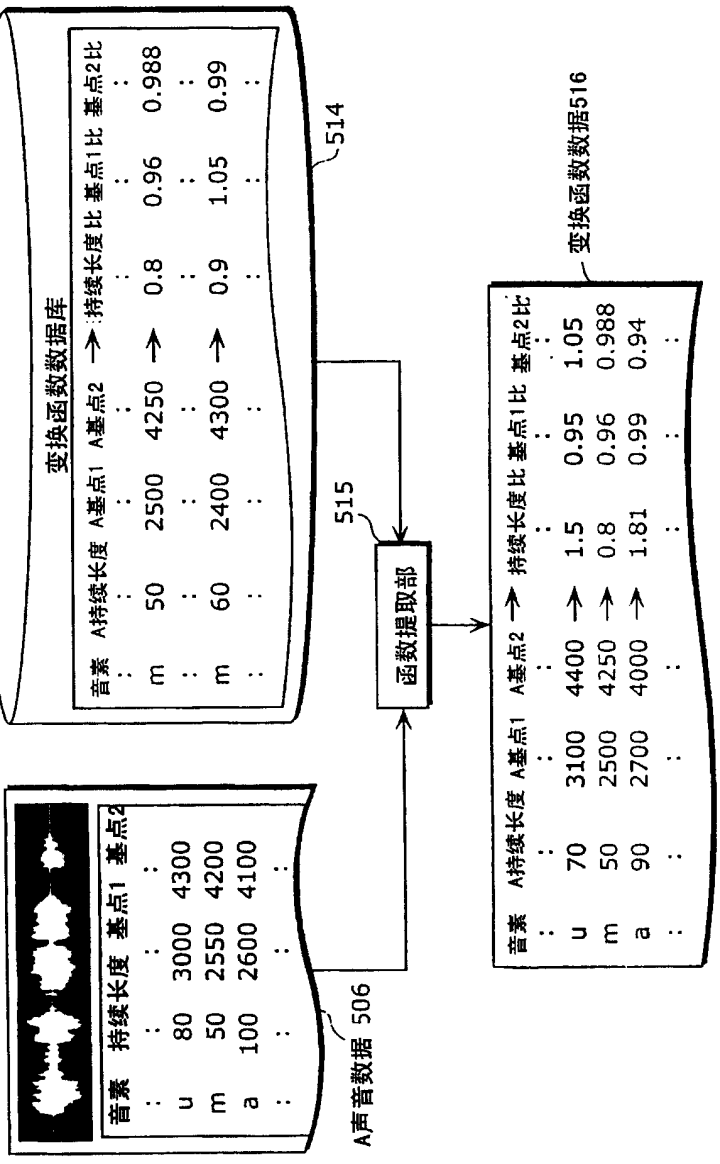

变换函数数据库
514
音素 A持续长度 A基点1 A基点2 → 持续长度比 基点1比 基点2比
m 50 2500 4250 → 0.8 0.96 0.988
m 60 2400 4300 → 0.9 1.05 0.99
515
函数提取部
A声音数据 506
音素 持续长度 基点1 基点2
u 80 3000 4300
m 50 2550 4200
a 100 2600 4100
音素 A持续长度 A基点1 A基点2 → 持续长度比 基点1比 基点2比
u 70 3100 4400 → 1.5 0.95 1.05
m 50 2500 4250 → 0.8 0.96 0.988
a 90 2700 4000 → 1.81 0.99 0.94
变换函数数据516

图27

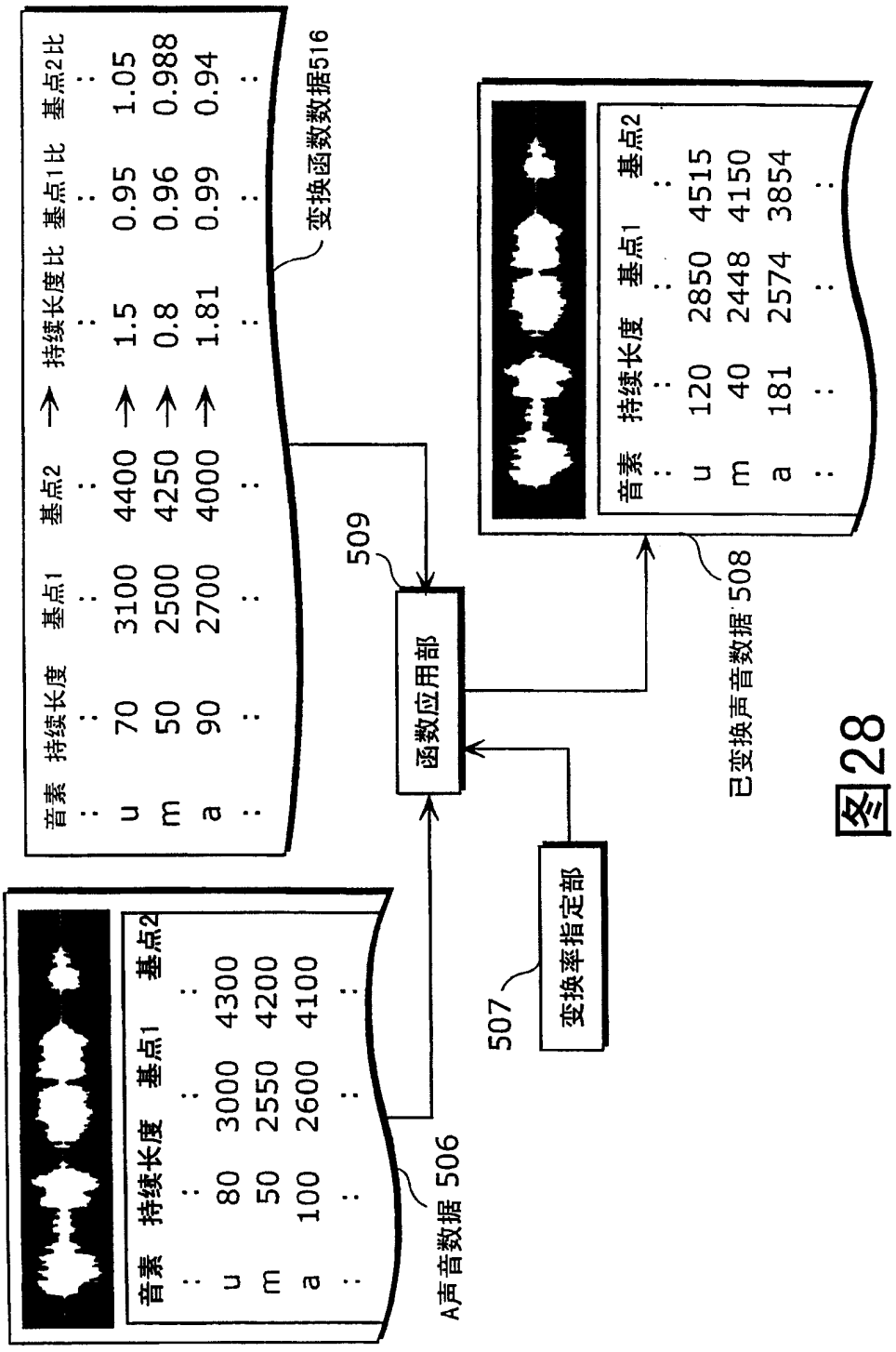

音素 持续长度 基点1 基点2
u 80 3000 4300
m 50 2550 4200
a 100 2600 4100
A声音数据 506
音素 持续长度 基点1 基点2 → 持续长度比 基点1比 基点2比
u 70 3100 4400 → 1.5 0.95 1.05
m 50 2500 4250 → 0.8 0.96 0.988
a 90 2700 4000 → 1.81 0.99 0.94
变换函数数据516
509
函数应用部
507
变换率指定部
音素 持续长度 基点1 基点2
u 120 2850 4515
m 40 2448 4150
a 181 2574 3854
已变换声音数据 508

图28

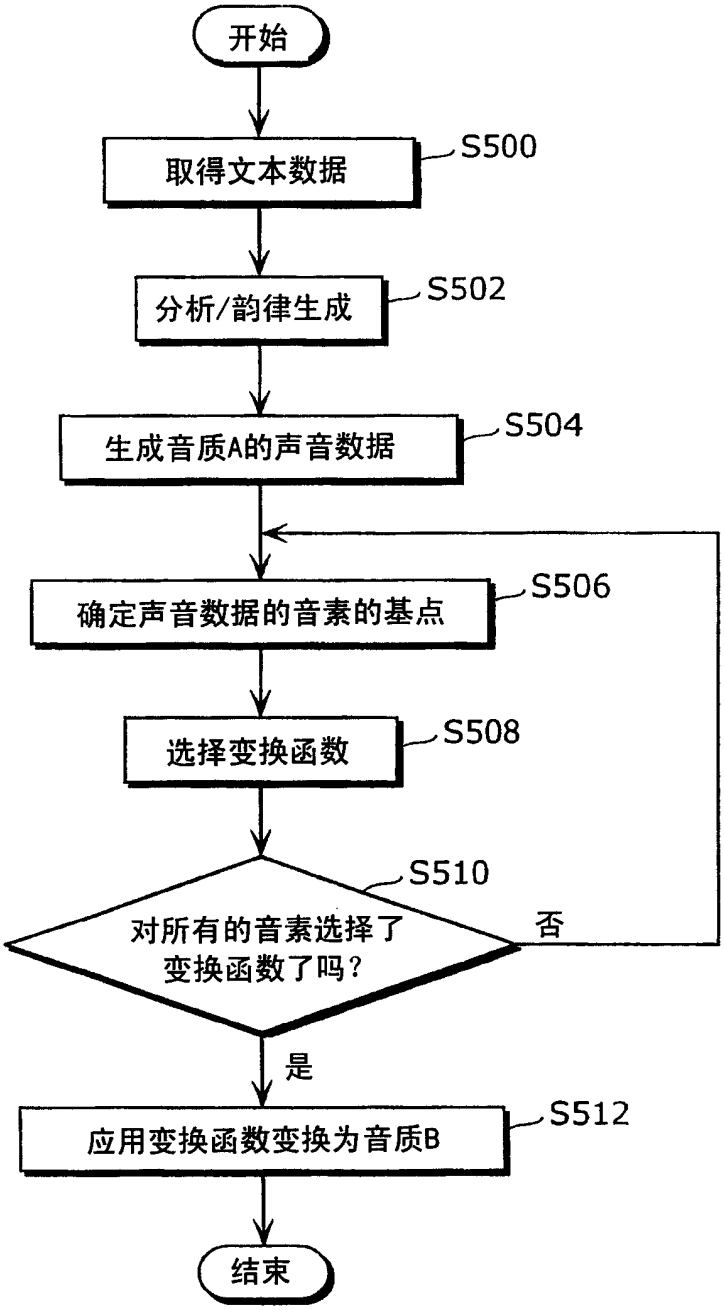

开始
取得文本数据
S500
分析/韵律生成
S502
生成音质A的声音数据
S504
确定声音数据的音素的基点
S506
选择变换函数
S508
对所有的音素选择了变换函数了吗？
S510
否
是
应用变换函数变换为音质B
S512
结束

图29

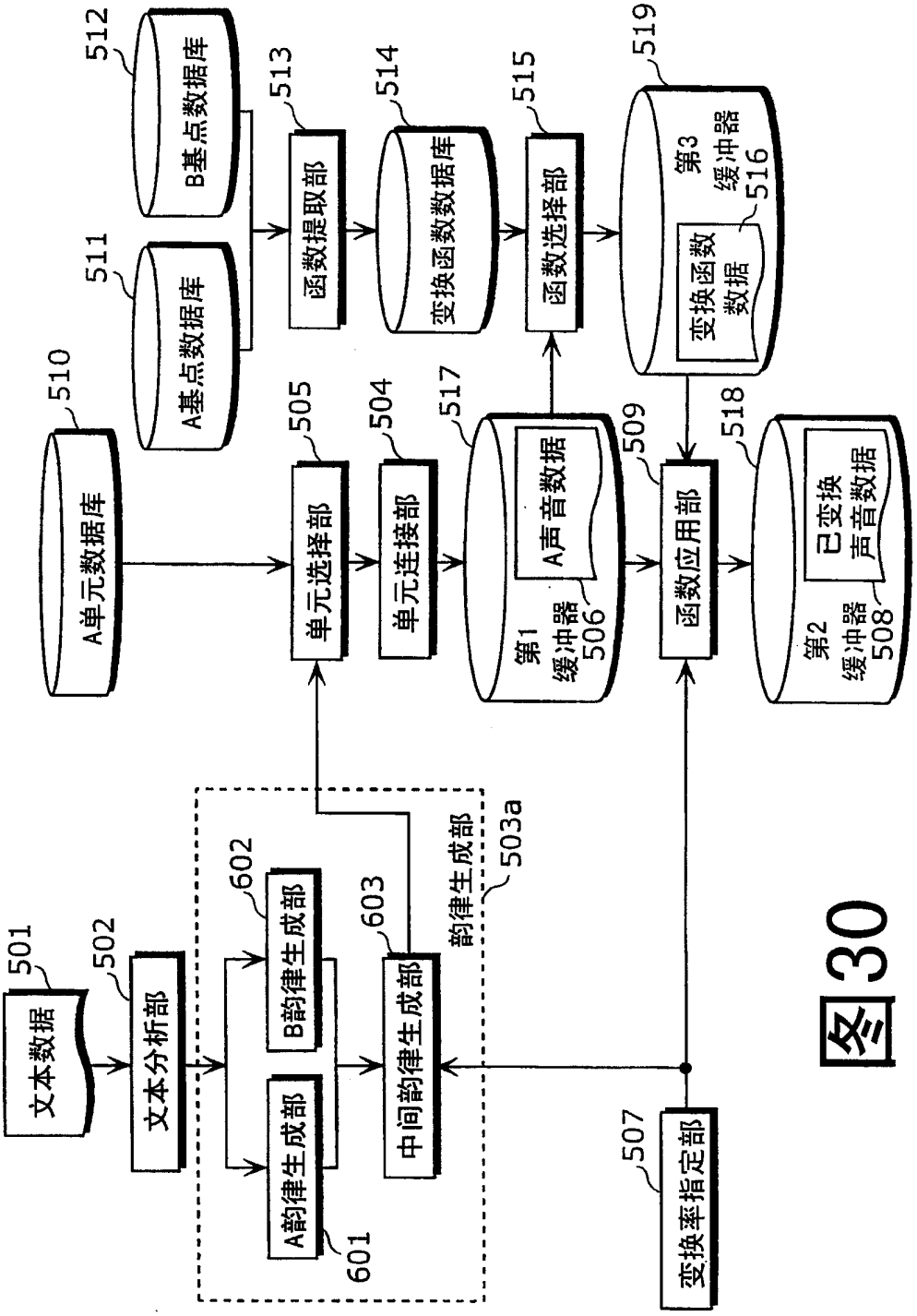

文本数据
501
文本分析部
502
韵律生成部
503a
A韵律生成部
601
B韵律生成部
602
中间韵律生成部
603
单元选择部
505
单元连接部
504
A单元数据库
510
A基点数据库
511
B基点数据库
512
函数提取部
513
变换函数数据库
514
函数选择部
515
第3缓冲器
516
变换函数数据
第1缓冲器
506
A声音数据
517
函数应用部
509
第2缓冲器
508
已变换声音数据
518
变换率指定部
507
519

图30

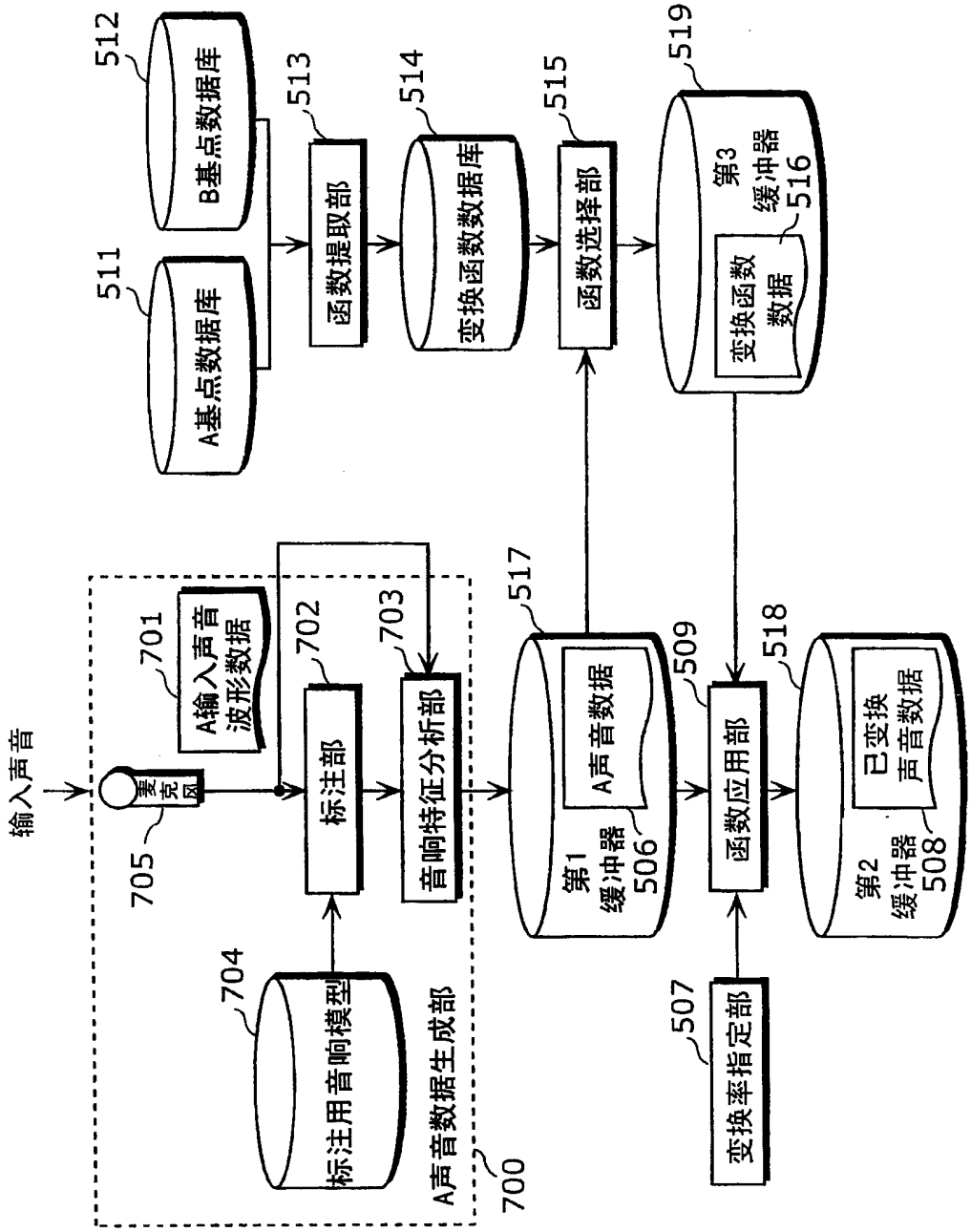

输入声音
麦克风
705
A输入声音波形数据
701
标注部
702
音响特征分析部
703
标注用音响模型
704
A声音数据生成部
700
A基点数据库
511
B基点数据库
512
函数提取部
513
变换函数数据库
514
函数选择部
515
第3缓冲器
516
变换函数数据
519
第1缓冲器
506
A声音数据
517
函数应用部
509
变换率指定部
507
第2缓冲器
508
已变换声音数据
518

图31