



# [12] 发明专利说明书

[21] ZL 专利号 00800399.8

[45] 授权公告日 2004 年 9 月 22 日

[11] 授权公告号 CN 1168068C

[22] 申请日 2000.3.27 [21] 申请号 00800399.8

[30] 优先权

[32] 1999. 3. 25 [33] JP [31] 81124/1999

[32] 1999. 7. 19 [33] JP [31] 204167/1999

[86] 国际申请 PCT/JP2000/001870 2000. 3. 27

[87] 国际公布 WO2000/058943 日 2000. 10. 5

[85] 进入国家阶段日期 2000. 11. 24

[71] 专利权人 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 松井谦二 釜井孝浩

山上胜义

审查员 刘红梅

[74] 专利代理机构 中国国际贸易促进委员会专利

商标事务所

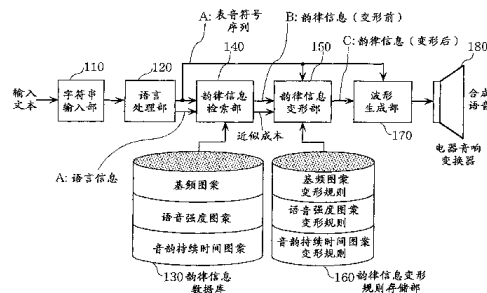
代理人 吴丽丽

权利要求书 6 页 说明书 23 页 附图 13 页

[54] 发明名称 语音合成系统与语音合成方法

[57] 摘要

韵律信息数据库(130)中存储有从实际语音提取出的韵律信息以及与之相对应的音素序列与重音位置。韵律信息检索部(140)根据基于输入文本等的语言处理部(120)的输出即音素序列等,从数据库(130)检索近似成本最小的韵律信息,韵律信息变形部(150)据近似成本,根据韵律信息变形规则存储部(160)中存储的变形规则使检索的韵律信息变形,使电气音响变换器(180)发出合成语音。由此,即使数据库(130)内设有与输入文本相对应的发音内容,也能发出同样音质的合成语音。



1.一种语音合成系统，用于根据表明合成语音的合成语音信息而输出合成语音其特征在于，包括有：

对应于成为检索关键词的关键词信息，存储用于语音合成的韵律信息的数据库；

用近似成本表示上述合成语音信息与上述关键词信息的一致程度，根据用该近似成本表示的一致程度，检索上述韵律信息的检索装置；

将对应于近似成本的变形程度作为变形规则存储的韵律信息变形规则装置；

根据上述合成语音信息与上述关键词信息的一致程度以及上述韵律信息变形规则装置存储的变形规则，对由上述检索装置所检索的韵律信息施行变形的变形装置；

根据上述合成语音信息以及上述变形装置所变形的韵律信息，输出合成语音的合成装置。

2.权利要求 1 所述的语音合成系统，其特征在于所述合成语音信息与所述关键词信息分别含有表示合成语音的语音属性的表音符号序列。

3.权利要求 2 所述的语音合成系统，其特征在于所述合成语音信息与所述关键词信息还分别包括表示合成语音的语言属性的语言信息。

4.权利要求 2 所述的语音合成系统，其特征在于所述表音符号序列至少包括合成语音的音韵序列、重音位置、以及实质性地表示停顿的有无或停顿时间长短之中的某个。

5.权利要求 3 所述的语音合成系统，其特征在于所述语言信息至少包括合成的语音的语法信息以及意义信息两者中的某个。

6.权利要求 3 所述的语音合成系统，其特征在于此系统还设有分析输入上述语音合成系统的文本信息，而生成上述表音符号序列和上

述语言信息的语言处理装置。

7.权利要求 1 所述的语音合成系统，其特征在于所述合成语音信息与所述关键词信息分别在实质上包含有表示合成的语音中各音韵所属音韵类别的音韵类别序列。

8.权利要求 7 所述的语音合成系统，其特征在于此系统还设有将输入上述语音合成系统的与上述合成语音信息相对应的信息以及存储于所述数据库中的与所述关键词信息相对应的信息这两者中的至少某个，变换为音韵类别序列的变换装置。

9.权利要求 7 所述的语音合成系统，其特征在于所述音韵类别至少应用音韵调音方式、调音位置以及持续时间长短之中的某个将音韵分组。

10.权利要求 7 所述的语音合成系统，其特征在于所述音韵类别用统计方法将韵律图案分组，而为了最好地反映韵律图案的组，用统计方法将音韵分组。

11.权利要求 10 所述的语音合成系统，其特征在于所述统计方法是多变量分析法。

12.权利要求 7 所述的语音合成系统，其特征在于所述音韵类别是根据音韵相互间异听表用统计方法决定的音韵间的距离将音韵分成的组。

13.权利要求 12 所述的语音合成系统，其特征在于所述统计方法是多变量分析法。

14.权利要求 7 所述的语音合成系统，其特征在于所述音韵类别是根据音韵物理特性的相似度将音韵分成的组。

15.权利要求 14 所述的语音合成系统，其特征在于所述物理特性至少是音韵的基频、强度、持续时间与音频频谱之中的某个。

16.权利要求 1 所述的语音合成系统，其特征在于，在存储于上述数据库中的上述韵律信息中包含有从同一实际语音中提取出的韵律特征。

17.权利要求 16 所述的语音合成系统，其特征在于，在示明有上

述韵律特征的信息中至少包含有：示明基频随时间变化的基频图案；示明语音强度随时间变化的语音强度图案；示明各音韵的持续时间的音频持续时间图案；以及示明停顿的有无或停顿持续时间这两者中之一的停顿信息。

18.权利要求 1 所述的语音合成系统，其特征在于，所述数据库按照各韵律控制单位存储上述韵律信息、

19.权利要求 18 所述的语音合成系统，其特征在于，上述韵律控制单位是加重字句、由 1 个以上加重文节构成的短语、文节、由 1 个以上文节构成的短语、单词、由 1 个以上单词构成的短语、重读字句以及由 1 个以上重读字句构成的短语之中的某个。

20.权利要求 1 所述的语音合成系统，其特征在于，上述合成语音信息以及上述关键词信息分别包含决定合成语音的要素即多种语音指标信息，而上述合成语音信息和关键词信息的一致程度则是由上述合成语音信息中各语音指标信息与上述关键词信息中各语音指标信息的一致程度分别加权合成的结果。

21.权利要求 20 所述的语音合成系统，其特征在于，上述语音指标信息至少包括对合成的语音的音频率列、重音位置、停顿的有无或停顿的持续时间以及表示语言属性的语言信息之中的某个作实质性表示的信息。

22.权利要求 21 所述的语音合成系统，其特征在于，上述语音指标信息包含有对合成的语音的音韵序列作实质性显示的信息，而上述合成语音信息中各语音指标信息与上述关键词信息中各语音指标信息的一致程度包含上述各音韵的音响特性长度的类似程度。

23.权利要求 20 的语音合成系统，其特征在于，上述语音指标信息实质上包含有示明合成的语音的各音韵所属音韵类别的音韵类别序列。

24.权利要求 23 的语音合成系统，其特征在于，上述合成语音信息中各语音指标信息与上述关键词信息中各语音指标信息的一致程度包含上述各音韵的音韵类别的相似程度。

25.权利要求 20 的语音合成系统，其特征在于，上述韵律信息包含有表征合成的语音的多种韵律特征信息。

26.权利要求 25 的语音合成系统，其特征在于，上述多种韵律特征信息集成组存储于所述数据库中。

27.权利要求 26 的语音合成系统，其特征在于，上述集成组的多种韵律特征信息都是由同一实际语音中提取出的。

28.权利要求 25 所述的语音合成系统，其特征在于，上述韵律的特征信息至少包含示明基频随时间变化的基频图案、示明语音强度随时间变化的语音强度图案、示明各音韵持续时间的音韵持续时间图案以及示明停顿有无或停顿持续时间的停顿信息之中的某个。

29.权利要求 28 所述的语音合成系统，其特征在于，上述音韵持续时间图案至少包含音素持续时间图案、拍的持续时间图案以及音节持续时间图案之中的某个。

30.权利要求 25 所述的语音合成系统，其特征在于，上述各种韵律特征信息分别根据通过不同的上述加权得到的上述合成语音信息与关键词信息的一致程度进行检索与变形。

31.权利要求 20 所述的语音合成系统，其特征在于，由上述检索装置进行的上述韵律的检索以及由上述变形装置进行的上述韵律信息的变形分别根据通过不同的上述加权得到的上述合成语音信息与关键词信息的一致程度进行。

32.权利要求 20 所述的语音合成系统，其特征在于，由上述检索装置进行上述韵律信息的检索与由上述变形装置进行的上述韵律信息的变形，分别根据通过相同的上述加权得到的上述合成语音信息与关键词信息的一致程度进行。

33.权利要求 1 所述的语音合成系统，其特征在于，上述变形装置根据至少是各音素、各拍、各音节、上述合成装置中各语音波形的生成单位以及各音韵之中的一个的一致程度，进行由上述检索装置检索的上述韵律信息的变形。

34.权利要求 33 所述的语音合成系统，其特征在于，上述各音素、

各拍、各音节、上述合成装置中各语音波形的生成单位以及各音韵中任何一个的一致度至少是基于根据音响特性的距离、根据调音方式与调音位置以及持续时间之中的某个求得的距离、以及基于收听试验求得的异听表的距离之中的某个所设定的。

35.权利要求 34 所述的语音合成系统，其特征在于，上述音响特性至少是基频、强度、持续时间与声频频谱中的某个。

36.权利要求 1 所述的语音合成系统，其特征在于，上述数据库对多种语言存储了上述关键词信息与韵律信息。

37.一种语音合成方法，它根据示明合成语音的合成语音信息，输出合成语音，其特征在于，

预先将对应于近似成本的变形程度作为变形规则存储在韵律信息变形规则装置中；

对应于成为检索关键词的关键词信息，从存储有语音合成用的韵律信息的数据库中，用近似成本表示上述合成语音信息与上述关键词信息的一致程度，对应于由该近似成本表示的一致程度，检索上述韵律信息；

根据上述合成语音信息与上述关键词信息的一致程度以及上述韵律信息变形规则装置中存储的变形规则，对上述检索装置所检索的韵律信息施行变形；

根据上述合成语音信息与上述变形装置所变形的韵律信息，输出合成语音。

38.权利要求 37 所述的语音合成方法，其特征在于，上述合成语音信息以及上述关键词信息分别包含决定合成语音的要素即多种语音指标信息，而上述合成语音信息和关键词信息的一致程度则是由上述合成语音信息中各语音指标信息与上述关键词信息中各语音指标信息的一致程度分别加权合成的结果。

39.权利要求 38 所述的语音合成方法，其特征在于：上述韵律信息包含有表征合成的语音的多种韵律特征信息。

40.权利要求 39 所述的语音合成方法，其特征在于，上述各种的

韵律特征信息分别根据由不同的上述加权得到的合成语音信息和关键词信息的一致程度进行检索与变形。

41.权利要求 38 所述的语音合成方法，其特征在于，由上述检索装置进行的上述韵律信息的检索以及由上述变形装置进行的上述韵律信息的变形，分别根据不同的上述加权所得的上述合成语音信息和关键词信息的一致程度进行。

42.权利要求 38 所述的语音合成方法，其特征在于，由上述检索装置进行的上述韵律信息的检索与由上述变形装置进行的上述韵律信息的变形，分别根据同一的由上述加权所得的上述合成语音合成信息与关键词信息的一致程度进行。

43.一种语音合成系统，它把输入的文本变换成合成语音而输出，此系统的特征在于它包括有：

分析上述输入的文本而输出表音符号序列与语音信息的语音处理装置；

将由实际语音提取出的韵律的特征量与对应于合成的语音的表音符号序列与语言信息相对应地加以存储的韵律信息数据库；

与上述语言处理装置输出的上述表音符号序列与上述语言信息组成的检索项目的至少一部分相对应，检索存储于上述韵律信息数据库中的上述韵律的特征量的检索装置；

用近似成本表示上述检索项目与上述韵律信息数据库的存储内容的一致程度，将与该近似成本表示的一致程度对应的变形程度作为变形规则进行存储的韵律信息变形规则装置；

根据上述检索项目与上述韵律信息数据库的存储内容的一致程度，将上述韵律信息数据库检索、选择的韵律特征量依存储在上述变形规则装置中对应于由上述近似成本表示的一致程度的变形规则变形的韵律变形装置；

根据上述韵律变形装置输出的上述韵律特征量和上述语言处理装置输出的上述表音符号序列，生成语音波形的波形生成装置。

## 语音合成系统与语音合成方法

### 技术领域

本发明涉及将任意输入的文本或输入的表音符号序列等，变换成合成语音输出的语音合成系统。

### 背景技术

近年来，在家电制品、汽车驾驶导向系统、手机等种种电子设备中，为了用语音表示设备的状态、操作等指示、应答消息等这样一类信息，多采用合成语音。此外，在个人计算机等之中也正在采用由语音接口进行的操作和由光学字符识别（OCR）进行字符识别结果的确认等。

作为进行上述这种语音合成的方法有首先存储音频数据然后使其再生的方法。这类方法多用于将有限的消息等以语音表示的场合中，采用这种方法，为了能发出任意的语音就必须有大容量的存储装置，成本高的结果限制了其用途。

另一方面，作为以较廉价的结构来发出任意语音的方法，则有根据输入的文本或表音符号序列的排列等，应用预定的音频数据生成规则来生成音频数据。但是采用这类音频数据，难以对各种各样的表示发出自然的语音。

例如据特开平 8-87297 号公报中所示，已知有应用数据库的语音信息检索进行合成语音的生成同时并用由合成语音生成规则进行合成语音的生成的语音合成系统。在这种系统对应的设备中，具体如图 13 所示，包括有字符串输入部 910、将通过分析实际语音提取出的音频特征量和与之对应的发声内容进行存储的语音信息数据库 920、检索语音信息数据库 920 的语音信息检索部 930、生成语音波形的合成语音生成部 940、包含有根据输入文本或输入表音符号序列生成音频特征量时的规则的合成语音生成规则 950、电声变换器 960。这种语音合

成系统当文本或表音符号序列输入字符串输入部 910 后，语音信息检索部 930 便从语音信息数据库 920 中检索与输入文本或输入表音符号序列相一致的发声内容的语音信息。当存在一致的发声内容时，便把对应的语音信息传递给合成语音生成部 940。当不存在一致的发声内容时，语音信息检索部 930 便把输入文本或输入表音符号序列原样地传送结合成语音生成部 940。合成语音生成部 940 在输入有检索出的语音信息时即据此生成合成语音，而当输入的是输入文本或输入表音符号序列时，则根据此以及合成语音生成规则 950，在生成音频特征量后生成合成语音。

如上所述，通过应用语音信息的检索和合成语音生成规则，能将任意输入的文本等变换为合成语音输出，与此同时，能使一部分语音（检索命中时）发出自然语音。

但在上述现有的语音合成系统中，在检索命中和未命中时，即语音信息数据库内存在和不存在与输入文本等相对应的发声内容时，音质的相差很大，通过将这样的音质不同的语音结合到一起，则会发生明显地不自然的问题。此外，由于仅仅是根据输入表音符号序列与存储的发声内容有无一致来进行语音信息数据库 920 的检索。当存在一致的发声内容时即不考虑语句的结构等便由可检索出的语音信息进行语言合成，也有成为颇不自然的合成语音的问题。

试考虑一具体例子，例如在语言合成“大阪に住んでいる私は松下です”这样的语句时，当专有各词“松下”不存在于数据库中时，便只有这个部分成为机械的合成语音，当通过利用作为语句末尾发声内容存储的“大阪に住んでいる”语音信息时，便常会成为将“大阪に住んでいる”、“私は松下です”这样两个语句不自然结合到一起的合成语音。

有鉴于上述种种问题，本发明的目的在于提供这样的语音合成系统，它能根据任意的输入文本发出自然合成的语音，特别是在即使语音信息（韵律信息）数据内不存在输入文本等所对应的发声内容时，也能以同样的音质发出合成语音。

发明内容

为实现上述目的，技术方案 1~6 的发明的特征在于，一种语音合成系统，用于根据表明合成语音的合成语音信息而输出合成语音其特征在于，包括有：对应于成为检索关键词的关键词信息，存储用于语音合成的韵律信息的数据库；用近似成本表示上述合成语音信息与上述关键词信息的一致程度，根据用该近似成本表示的一致程度，检索上述韵律信息的检索装置；将对应于近似成本的变形程度作为变形规则存储的韵律信息变形规则装置；根据上述合成语音信息与上述关键词信息的一致程度以及上述韵律信息变形规则装置存储的变形规则，对由上述检索装置所检索的韵律信息施行变形的变形装置；根据上述合成语音信息以及上述变形装置所变形的韵律信息，输出合成语音的合成装置。

上述合成语音信息以及上述关键词信息也可分别包括表示合成语音的语音属性的表音符号序列，或还包括有表示合成语音的语言属性的语音信息，上述表音符号序列可以至少包括合成的语音的音韵序列、重音位置以及对停顿的有无或停顿时间两者的某个作实质性表示的信息。此外，上述语言信息至少可包括合成的语音的语法信息以及意义信息两者中的某个。

上述系统的特征还在于包括有语言处理装置，它通过分析输入上述语音合成系统的文本信息，生成前述的表音符号序列以及前述的语言信息。

这样，即使数据库中没有存储合成语音信息与关键词信息完全一致的韵律信息，由于可根据类似的韵律信息进行语音的合成，故能对于任意的语音发出比较适当而且没有忽高忽低的自然语音。相反，还能不损害合成语音的自然性和减少数据库的存储容量。再有，在采用上述那种类似的韵律信息时，由于能根据其类似程度使韵律信息变形，就能发出更确切的合成语言。

技术方案 7~15 的发明的特征在于，在技术方案 1 的语音合成系统中，上述合成语音信息以及上述关键词信息都在实质上包含表示合成的语音中各音韵所属音韵类别的音韵类别序列。

上述系统的特征还在于包括有变换装置，它把输入到上述语音合成系统中的上述合成语音信息所对应的信息以及存储于上述数据库的所述关键词信息对应的信息这两者中至少某个变换到音韵分类序列中。

上述音韵类别至少可以是：利用音韵的调音方式、调音位置以及持续时间中的某个将音韵分组的结果；

应用统计方法使韵律图案分组化，为了更良好地反映韵律图案的组别再采用多变量分析等统计方法将音韵分组化的结果；

由音韵相互间的异听表，根据多变量分析等统计方法所确定的音韵间的距离，将音韵分组化的结果；

或是根据音韵的基频、强度、持续时间或声频频谱等物理特性的相似度，将音韵分组化的结果。

于是，在韵律信息的检索中，即使在音素序列不一致时，对于各音素的音韵分类相一致的情形，纵使滥用了音韵，在多数情形下也能发出自然的合成语音。

技术方案 16 的发明是使技术方案 1 的语音合成系统具备下述特征，即在存储于上述数据库中的上述韵律信息中包含有从同一实际语音中提取出的韵律特征。

技术方案 17 的发明是使技术方案 16 的语音合成系统具备下述特征，即在示明有上述韵律特征的信息中至少包含有：示明基频随时间变化的基频图案；示明语音强度随时间变化的语音强图图案；示明各音韵的持续时间的音韵持续时间图案；以及示明停顿的有无或停顿持续时间这两者中之一的停顿信息。

技术方案 18 的发明是使技术方案 17 的语音合成系统具备下述特征，即所述数据库是以各韵律控制单位存储上述韵律信息。

技术方案 19 的发明是使技术方案 18 的语音合成系统具备下述特征：上述韵律控制单位是加重字句、由 1 个以上加重字句构成的短语、文节、由 1 个以上文节构成的短语、单词、由 1 个以上单词构成的短语、重读字句以及由 1 个以上重读字句构成的短语等之中的某个。

由此就容易发出适当的自然的合成语音。

技术方案 20 的发明是使技术方案 1 的语音合成系统具备下述特征：上述合成语音信息以及上述关键词信息分别包含决定合成语音的要素即多种语音指标信息，而上述合成语音信息和关键词信息的一致程度则是由上述合成语音信息中各语音指标信息与上述关键词信息中各语音指标信息的一致程度分别加权合成的结果。

技术方案 21 的发明是使技术方案 20 的语音合成系统具备下述特征：上述语音指标信息至少包括对合成的语音的音韵序列、重音位置、停顿的有无或停顿的持续时间以及表示语言属性的语言信息等之中的某个作实质性表示的信息。

技术方案 22 的发明是使技术方案 21 的语音合成系统具备下述特征：上述语音指标信息包含有对合成的语音的音韵序列作实质性显示的信息，而上述合成语音信息中各语音指标信息与上述关键词信息中各语音指标信息的一致程度包含上述各音韵的特性长度的类似程度。

技术方案 23 的发明是使技术方案 20 的语音合成系统具备下述特征：上述语音指标信息实质上包含有示明合成的语音的各音韵所属音韵类别的音韵类别序列。

技术方案 24 的发明是使权利标 23 的语音合成系统具备下述特征：上述合成语音信息中各语音指标信息与上述关键词信息中各语音指标信息的一致程度包含上述各音韵的音韵类别的相似程度。

由此就易于进行适当的韵律信息的检索与变形。

技术方案 25 的发明是使技术方案 20 的语音合成系统具备下述特征：上述韵律信息包含有表征合成的语音的多种韵律特征信息。

技术方案 26 的发明是使技术方案 25 的语音合成系统具备下述特征：上述多种韵律特征信息集成组存储于所述数据库中。

技术方案 27 的发明是使技术方案 26 的语音合成系统具备下述特征：上述集成组的多种韵律特征信息都是由同一实际语音中提取出的。

技术方案 28 的发明是使技术方案 25 的语音合成系统具备下述特

征：上述韵律的特征信息至少包含示明基频随时间变化的基频图案、示明语音强度随时间变化的语音强度图案、示明各音韵持续时间的音韵持续时间图案以及示明停顿有无或停顿持续时间的停顿信息等之中的某个。

技术方案 29 的发明是使技术方案 28 的语音合成系统具备下述特征：上述音韵持续时间图案至少包含音素持续时间图案、拍的持续时间图案以及音节持续时间图案等之中的某个。

技术方案 30 的发明是使技术方案 25 的语音合成系统具备下述特征：上述各种韵律特征信息分别根据通过不同的上述加权得到的上述合成语音信息与关键词信息的一致程度进行检索与变形。

技术方案 31 的发明是使技术方案 20 的语音合成系统具备下述特征：由上述检索装置进行的上述韵律的检索以及由上述变形装置进行的上述韵律信息的变形分别根据通过不同的上述加权得到的上述合成语音信息与关键词信息的一致程度进行。

技术方案 32 的发明是使技术方案 20 的语音合成系统具备下述特征：由上述检索装置进行上述韵律信息的检索与由上述变形装置进行的上述韵律信息的变形，分别根据通过相同的上述加权得到的上述合成语音信息与关键词信息的一致程度进行。

技术方案 33 的发明是使技术方案 1 的语音合成系统具备下述特征：上述变形装置根据至少是各音素、各拍、各音节、上述合成装置中各语音波形的生成单位以及各音韵等之中的一个的一致程度，进行由上述检索装置检索的上述韵律信息的变形。

技术方案 34 的发明是使技术方案 33 的语音合成系统具备下述特征：上述各音素、各拍、各音节、上述合成装置中各语音波形的生成单位以及各音韵中任何一个的一致度至少是基于根据音响特性的距离、据调音方式与调音位置以及持续时间之中的某个求得的距离、以及由收听试验求得的异听表而有的距离等之中的某个所设定的。

由此就容易进行各切的变形。

技术方案 35 的发明是使技术方案 34 的语音合成系统具备下述特

征：上述音响特性至少是基频、强度、持续时间与声频频谱中的某个。

技术方案 36 的发明是使技术方案 1 的语音合成系统具备下述特征：上述数据库对多种语言存储了上述关键词信息与韵律信息。

由此能容易发出包括多种语言在内的合成语音。

技术方案 37 的发明的特征在于，一种语音合成方法，它根据示明合成语音的合成语音信息，输出合成语音，其特征在于，预先将对应于近似成本的变形程度作为变形规则存储在韵律信息变形规则装置中；对应于成为检索关键词的关键词信息，从存储有语音合成用的韵律信息的数据库中，用近似成本表示上述合成语音信息与上述关键词信息的一致程度，对应于由该近似成本表示的一致程度，检索上述韵律信息；根据上述合成语音信息与上述关键词信息的一致程度以及上述韵律信息变形规则装置中存储的变形规则，对上述检索装置所检索的韵律信息施行变形；根据上述合成语音信息与上述变形装置所变形的韵律信息，输出合成语音。

技术方案 38 的发明是使技术方案 37 的语音合成方法具备下述特征：上述合成语音信息以及上述关键词信息分别包含决定合成语音的要素即多种语音指标信息，而上述合成语音信息和关键词信息的一致程度则是由上述合成语音信息中各语音指标信息与上述关键词信息中各语音指标信息的一致程度分别加权合成的结果。

技术方案 39 的发明是使技术方案 38 的语音合成方法具备下述特征：上述韵律信息包含有表征合成的语音的多种韵律特征信息。

技术方案 40 的发明是使技术方案 39 的语音合成方法具备下述特征：上述各种的韵律特征信息分别根据由不同的上述加权得到的合成语音信息和关键词信息的一致程度进行检索与变形。

技术方案 41 的发明是使技术方案 38 的语音合成方法具备下述特征：由上述检索装置进行的上述韵律信息的检索以及由上述变形装置进行的上述韵律信息的变形，分别根据不同的上述加权所得的上述合成语音信息和关键词信息的一致程度进行。

技术方案 42 的发明是使技术方案 38 的语音合成方法具备下述特

征：由上述检索装置进行的上述韵律信息的检索与由上述变形装置进行的上述韵律信息的变形，分别根据同一的由上述加权所得的上述合成语音合成信息与关键词信息的一致程度进行。

由此仍然会有，即使数据库中未存储合成语音信息与关键词信息完全一致的韵律信息，但由于能根据类似的韵律信息进行语音合成，故可相对于任意的语音，发出比较适当且无忽高忽低的自然语言。相反，还能不损害合成语音的自然性而可减小数据库的存储容量。此外，在与上述类似采用韵律信息的情形，由于是根据其类似程度使韵律信息变形，便能发出更适当的合成语音。

技术方案 43 的发明是使把输入的文本变换成合成语音而输出的语音合成系统具备下述特征，即使此系统包括：分析输入的文本而输出表音符号序列与语言信息的语言处理装置；将由实际语音求出的韵律的特征量与对应于合成的语音的表音符号序列与语言信息相对应地加以存储的韵律信息数据库；与上述语言处理装置输出的上述表音符号序列与上述语言信息组成的检索项目的至少一部分相对应，检索存储于上述韵律信息数据库中的上述韵律的特征量的检索装置；根据上述检索项目与上述韵律信息数据库的存储内容的一致程度，将上述韵律信息数据库检索、选择的韵律特征量依所定规则变形的韵律变形装置；根据上述韵律变形装置输出的上述韵律特征量和上述语言处理装置输出的上述表音符号序列，生成语音波形的波形生成装置。

由此仍然能相对于任意的输入文本发出比较适当而无或高或低的自然语言。

#### 附图说明

图 1 是示明实施形式 1 的语音合成系统结构的功能框图。

图 2 是例示实施形式 1 的语音合成系统各部分信息的说明图。

图 3 是示明实施形式 1 的语音合成系统的韵律信息数据库的存储内容的说明图。

图 4 是例示基本频率图案的变形的说明图。

图 5 是例示韵律信息的变形的说明图。

图 6 是示明实施形式 2 的语音合成系统结构的功能框图。

图 7 是示明实施形式 2 的语音合成系统的韵律信息数据库的存储

内容的说明图。

图 8 是示明实施形式 3 的语音合成系统结构的功能框图。

图 9 是示明实施形式 4 的语音合成系统结构的功能框图。

图 10 是示明实施形式 4 的语音合成系统的韵律信息数据库的存储内容。

图 11 是示明实施形式 5 的语音合成系统结构的功能框图。

图 12 是音韵类别的说明图。

图 13 是示明先有的语音合成系统结构的功能框图。

具体实施方式

下面根据实施形式具体说明本发明的内容。

(实施形式 1)

图 1 是示明实施形式 1 的语音合成系统结构的功能框图。在图 1 中：

字符串输入部 110 是把输入的汉字假名交混的字符串与假名字符串等的文本等作为语音合成对象的信息。这种字符串输入部 110 具体地说例如可以采用键盘之类的输入装置等；

语言处理部 120 用于后述数据库检索等的前处理，它分析输入的文本，例如图 2 所示，对各个加重字句输出表音符号序列以及语言信息。在此，上述加重字句为方便起见取为用于语音合成的处理单位，大致相当于语法上的文节，例如 2 位以上的数字将各位数字分别设定为 1 个加重字句等，以适应语音合成处理而把输入文本加以区分的结果。此外，上述表音符号序列例如是根据字母数字符号组成的字符串来表示成为语音发声单位的音素与重音的位置等。再者，上述语音信息例如是示明加重字句的语法信息（词类等）以及意义信息（意义的属性等）。

韵律信息数据库 130，例如图 3 所示，是对于各个加重字句中从实际语音中对各加重字句抽出的韵律信息，与被检索关键词相对应地加以存储的结果，作为被检索关键词可以采用：

(a)音素序列，

(b)重音位置，

- (c) 拍数,
- (d) 加重句前后的停顿时间
- (e) 语法信息以及意义信息。

作为韵律信息可以采用

- (a) 基频图案,
- (b) 音强图案,
- (c) 音韵持续时间图案。

在此,为了发出自然的合成语言,上述各韵律信息最好是由同一实际语音中提取出。此外,上述拍数也可不预先存储于韵律信息数据库 130 中,而在每次检索时根据上述音素序列计数出。此外,上述加重语句前后的停顿时间在此同一图的例子中还同时表示了加重语句是在文首还是在文尾的信息。由此,即使在同一加重语句因文中的位置致使发音强度等不同的情形中,可在检索中区别而能合成适当的语音,但并不局限于此,也可以只包含停顿时间,还可以把表示文首、文尾的信息作为另外的被检索关键词。

韵律信息检索部 140 根据语言处理部 120 的输出,检索韵律信息数据库 130 的韵律信息再行输出。在此检索中进行所谓的模糊检索。具体地说,即使根据语言处理部 120 的输出的音素序列等的检索关键词与韵律信息数据库 130 中的被检索关键词并不完全一致,也可以把有某种程度一致的作为检索候补,由其中例如根据最低成本法,选择一致程度最高的(相当于检索关键词与被检索关键词的差的近似成本低的)。也就是,即使检索关键词与被检索关键词不完全一致,也能通过采用类似的加重语句的韵律信息,由生成规则生成韵律信息而发出自然的语音。

韵律信息变形部 150 根据韵律信息检索部 140 检索时的近似成本以及后述的韵律信息变形规则存储部 160 中存储的变形规则,将韵律信息检索部 140 所检索的韵律信息变形。具体地说,在韵律信息检索部 140 的检索中,当检索关键词与被检索关键词一致时,根据检索出的韵律信息可以进行最适当的语言合成,但当这两个关键

词并不完全一致时，如上所述，由于采用了类似的加重字句的韵律信息，这两个关键词的一致程度越低（近似成本越大），合成语音就有可能偏离适当的语音。在此，根据上述近似成本，对检索出的韵律信息施加预定的变形，由此即可发出更适当的合成语音。

韵律信息变形规则存储部 160 存储着用于对应于上述近似成本的韵律信息变形的变形规则。

波形生成部 170 根据语言处理部 120 输出的表音符号序列和韵律信息变形部 150 输出的韵律信息，合成语音波形而输出模拟语音信号。

电气音响变换器 180 例如是扬声器或头戴耳机等，将模拟语音信号变换为语音。

下面说明依上所述构成的语音合成系统的语音合成作业。

(1) 在拟变换为语音的文本被输入字符串输入部 110 后，语言处理部 120 分析输入的文本，解析各加重字句，输出图 2 所示的表音符号序列与语言信息。具体举例说明如下，在输入汉字假名交混的字符串时，应用图中未示明的汉字词典等一类的变换辞典，在分离加重字句的同时，变换读法，生成表示重音位置与停顿有无和停顿时间等的符号序列。在图 2 的表音符号序列的例子中，由字母数字符号表示下述信息。

- (a) 字母：音素（“N”表示拨音）
- (b) “,”：重音位置
- (c) “/”：加重句的区分
- (d) “C1”：无音区间
- (e) 数字：停顿时间

再有，图中虽然未示明，但它也表明了示明短语或句子分离的信息。表音符号序列的表示法也不限于以上所述，也可把表示音素序列或重音位置的数值等分别作为另外的信息输出。作为语言信息（语法信息、意义信息），除词类或意义等之外，还可包括活用形、系助词的受体的有无、一般句子中的重要度等，表示方法也不限于

此图中所示的“名词”与“连体形”等字符串，而是可以采用被编码的数字。

(2) 韵律信息检索部 140 根据语言处理部 120 输出的各个加重语句的表音符号序列和语言信息，检索韵律信息数据库 130 的韵律信息，检索出的韵律信息与以后详述的近似成本一起输出。更详细地说，在从语言处理部 120 输出上述列表表示的表音符号序列时，首先根据此表音符号序列求出音素序列、重音位置、拍数等作为检索关键词，去检索韵律信息数据库 130 中的韵律信息。在这种检索中，当韵律信息数据库 130 中存在与上述检索关键词完全一致的被检索关键词时，则可将与此被检索关键词相对应的韵律信息作为检索结果，但当其不存在时，则首先将有某种程度一致的（例如音素序列虽一致而意义信息不一致的，或音素序列虽不一致但重音与拍数一致）作为检索候补，从其中选择检索关键词与被检索关键词一致程度最高的，作为检索结果。

以上选择例如可用近似成本的最小成本法进行。具体地说，首先如下求近似成本 C。

(式 1)

$$C = a_1 \cdot D_1 + a_2 \cdot D_2 + a_3 \cdot D_3 + a_4 \cdot D_4 + a_5 \cdot D_5 + a_6 \cdot D_6 + a_7 \cdot D_7$$

上式中的  $a_1 \cdot D_1$  等的意义如下：

D1：音素序列中不一致的音素数，

D2：重音位置的差，

D3：拍数的差，

D4：在紧前方的停顿时间有无一致（是否在被检索关键词的范围内），

D5：在紧后方的停顿时间有无一致（是否在被检索关键词的范围内）

D6：语法信息一致的有无或一致程度，

D7：意义信息一致的有无或一致程度，

$a_1 \sim a_7$  则为上述 D1 ~ D7 的加权系数（上述 D1 ~ D7 是通过

统计方法与学习求得的与适当的韵律信息的选择相关的程度)

此外,作为 D1 ~ D7,也并不限于以上所述,若是表示的是检索关键词与被检索关键词的一致程度,则可以利用种种内容。例如对 D1 来说,也可以根据不一致的音素是否是相互类似的音素、不一致的音素位置、不一致的音素是否连续等而设定不同的值。对 D4、D5 而言,当停顿时间如图 3 所示的长、短、无等阶段表示时,可用 0、1 表示是否一致,可用数值表示阶差,而停顿时间长短以时间的数值表示时,也可应用时间的差数。对 D6、D7 来说,语法信息或意义信息的是否一致可用 0、1 表示,应用以检索关键词和被检索关键词为参数的表,也可以采用表示相应于这两者组合一致程度(例如各词与动词的一致程度低而助词与助词的一致程度高)的数值,或也可应用同义语词典求也意义相似的程度。

对各检索候补算出上述的近似成本,把近似成本最低的选作为检索结果,由此,即使韵律信息数据库 130 中没有存储检索键与被检索键完全一致的韵律信息,也可通过相似的韵律信息发出比较适当且自然的语音。

(3) 韵律信息变形部 150 根据韵律信息检索部 140 输出的近似成本,应用韵律信息变形规则存储部 160 中存储的规则,把作为韵律信息检索部 140 的检索结果而输出的韵律信息(基频图案、语言强度图案、音韵持续时间图案)变形。具体地说,例如在适用将基频图案的动态范围压缩的变形规则,就成为图 4 所示的基频图案变形。

对应于上述近似成本的变形具有下述意义。例如图 5 所示,当相对于输入文本“門真市”检索了“名古屋市”的韵律信息后,它们的音素序列虽然不同,但由于其他的检索项目一致(近似成本低),若将“名古屋市”的韵律信息原样地不变形地使用,也能进行适当的语音合成。再有,例如相对于“5分です”检索到“なるんです”时,为了获得“なるんです”的适当的合成语音,一般在考虑到词类的不同时,最好将“なるんです”的语音强度图案多少

有些减小，若是考虑文节信息（例如意义的重要程度）时，由于数字在多数情形下发音的强度大，就最好使“なるんてす”的语音强度图案作某种程度的增大，总的说来，最好将“なるんてす”的语音强度图案多少增大一些。这种综合的变形程度由于与近似成本有相关关系，通过将对应于近似成本的变形程度（变形倍率等）作为变形规则存储于韵律信息变形规则存储部 160 中，便可得到适当的合成语音。再有，韵律信息的变形不限于图 4 所示在历经的整个时间内作相同的变形，例如也可以主要在时间历程的中途附近通过使之变形等的变形图案，随着时间历程而使变形程度不同。作为上述变形规则的具体存储形式，可以把用于将近似成本变换为变形倍率的系数设定为变形规则，也可将近似成本作为参数，采用对应于变形倍率或变形图案的表。再有，作为变形用的近似成本，不限于与上述检索中所用近似成本相同，上面的（式 1）中，由于系数  $a_1 \sim a_7$  不同，可以去求得进行更适当变形的值，此外，对于基频图案、语音强度图案、音韵持续时间图案也可采用不同的值。例如，在（式 1）的各项可取负值的情形，即可把各项的绝对值的和作为检索用的近似成本（0 或正），也可把各项原样的值的和作为变形用的近似成本（可以为负）使用，等等。

（4）波形生成部 170 根据语言处理部 120 输出的表音符号序列和韵律信息变形部 150 变形的韵律信息，亦即根据音素序列与停顿时间、基频图案、语音强度图案以及音韵持续时间图案合成语音波形，输出模拟语音信号。根据这种模拟语音信号由电气音响的变换器 180 发出合成语音。

如上所述，即使在韵律信息数据库 130 没有存储检索关键词与被检索关键词完全一致的韵律信息，由于能根据相似的韵律信息进行语音合成，就可发出比较适当且无或高或低的自然语音。相反，还能无损于合成语音的自然性而减小韵律信息数据库 130 的存储容量。此外，如上所述，在采用相似的韵律信息情形，由于能根据其相似程度使韵律信息变形，可以发出较适当的合成语音。

## (实施形式2)

现以加重字句前后的停顿时间也作为韵律信息存储于韵律信息数据库中的语音合成系统为例,说明实施形式2的语音合成系统。在下述的实施形式中,具有与上述实施形式1等相同功能的结构部件附以相同或相应的标号而略去其详细说明。

图6为示明实施形式2的语音合成系统结构的功能框图。这一语音合成系统与实施形式1的语音合成系统相比有以下几点不同。

(a)语言处理部220与语言处理部120不同,输出不含有停顿信息的表音符号序列。

(b)韵律信息数据库230如图7所示与韵律信息数据库130不同,存储着不把停顿信息作为被检索关键词的韵律信息。实际上,与韵律信息数据库130相同应用相同的数据结构,在检索时也可把停顿时间作为韵律信息处理。

(c)韵律信息检索部240通过不含停顿信息的检索关键词、被检索关键词的核对进行检索,(在基频图案、语音强度图案、音韵持续时间图案之外还有)停顿信息也作为韵律信息输出。

(d)韵律信息变形部250根据近似成本使停顿信息也与基频图案等相同进行变形。

(e)韵律信息变形规则存储部260存储有基频图案变形规则等同时,还存储停顿时间变更规则。

如上所述,通过采用从韵律信息数据库230检索到的停顿信息,能够发出停顿时间较自然的合成语音,此外还能减轻语言处理部220中输入文本分析处理的负担。

与实施形式1相同,要是在检索时把语言处理部输出的停顿信息也用作检索关键词,则易于提高检索精度。此时,韵律信息数据库中另外存储或兼用用作被检索键的停顿信息和用作韵律信息的停顿信息。这样,在停顿信息从语言处理部输出的同时还存储于韵律信息数据库中的情形,采用哪种停顿信息来进行语音合成则可以根据语言处理部的分析精度和从韵律信息数据库检索停顿信

息的可靠性进行选择,也可以根据近似成本(检索结果的确切性)来确定选择哪种停顿信息。

(实施形式3)

下面以基频图案等分别基于另外的近似成本进行韵律信息的检索与变形的语音合成系统为例,来说明实施形式3的语音合成系统。

图8是示明实施形式3的语音合成系统结构的功能框图。此语音合成系统与上述实施形式1的语音合成系统相比有以下几点不同。

(a)取代韵律信息检索部140,设有基频图案检索部341、语音强度图案检索部342以及音韵持续时间图案检索部343。

(b)取代韵律信息变形部150,设有基频图案变形部351、语音强度图案变形部352以及音韵持续时间图案变形部353。

以上各检索部341~343和各变形部351~353分别用由以(式2)~(式4)求得的近似成本,独立地检索(选择检索候补)或变形基频图案、语音强度图案或音韵持续时间图案。

(式2)(基频图案的检索、变形)

$$C=b_1 \cdot D_1+b_2 \cdot D_2+b_3 \cdot D_3+b_4 \cdot D_4+b_5 \cdot D_5+b_6 \cdot D_6+b_7 \cdot D_7$$

(式3)(语音强度图案的检索、变形)

$$C=c_1 \cdot D_1+c_2 \cdot D_2+c_3 \cdot D_3+c_4 \cdot D_4+c_5 \cdot D_5+c_6 \cdot D_6+c_7 \cdot D_7$$

(式4)(音韵持续时间图案的检索、变形)

$$C=d_1 \cdot D_1+d_2 \cdot D_2+d_3 \cdot D_3+d_4 \cdot D_4+d_5 \cdot D_5+d_6 \cdot D_6+d_7 \cdot D_7$$

上述 $D_1 \sim D_7$ 虽与实施形式1的(式1)相同,但加权系数 $b_1 \sim b_7$ 、 $c_1 \sim c_7$ 、 $d_1 \sim d_7$ 则与(式1)的 $a_1 \sim a_7$ 不同,要分别采用进行适当的基频图案、语音强度图案或音韵持续时间图案的选择,由统计方法和学习求得的结果。具体地说,一般的基频图案当重音位置与拍数相同时大致类似,为此设定系数 $b_2$ 、 $b_3$ 比(式1)的系数 $a_2$ 、 $a_3$ 大。此外,语音强度图案对停顿的有无与停顿时间长短有很大的依赖性,故设定系数 $C_4$ 、 $C_5$ 比系数 $a_4$ 、 $5a_5$ 大。同

样, 由于音韵持待时间图案对音素序列的依赖性大, 故设定系数  $d1$  比系数  $a1$  大。

如上所述, 通过采用另外的近似成本独立地进行基频图样等的检索、变形, 就可进行均衡良好的检索与变形, 可根据各最适当的基频图样等进行语音合成。此外, 在韵律信息数据库 130 中不必将基频图案、语音强度图案与音韵持续时间图案成组地存储, 例如可以只存储各个图样的种类数量, 因而能够由较小存储容量的韵律信息数据库 130 发出音质良好的合成语音。

#### (实施形式 4)

现在说明实施形式 4 的语音合成系统。

图 9 是示明实施形式 4 的语音合成系统结构的功能框图。此语音合成系统主要具有下述特征。

(a) 与上述实施形式 1 ~ 3 不同, 韵律信息的检索与变形等的处理, 不是以加重字句单位而是以短语为单位进行。这里的短语也称之为短句或呼气段落等, 通常在发声时 (与有句点时相同) 被分离的 1 个或多个加重字句的集合。

(b) 与实施形式 2 相同, 设有存储了作为韵律信息的停顿信息的韵律信息数据库 430, 以及存储了基频图案变形规则与停顿时间变更规则的韵律信息变形规则存储部 460。它们如图 10 所示, 就韵律信息与变形规则等以短语单位存储这一点而言, 是与实施形式 2 的韵律信息数据库 230 和韵律信息变形规则存储部 260 不同。

(c) 与实施形式 3 相同, 韵律信息的检索与变形由基频图案等根据各个近似成本进行。此外, 停顿信息的检索与停顿时间的变更同样独立地进行。

(d) 韵律信息的变形与实施形式 1 ~ 3 相同, 根据近似成本进行, 同时还根据检索关键词与被检索关键词的音素序列中各音素的一致度 (一致的程度或有无) 进行, 而这是与实施形式 1 ~ 3 的不同处。

下面作更详细地说明。

语言处理部 420 与实施形式 1 的语言处理部 120 相同，分析字符串输入部 110 输入的文本，将各个加重字句分离后，以所定的加重字句的汇集结果短语为单位而输出表音符号序列和语言信息。

韵律信息数据库 430 如上所述以短语单位存储韵律信息，与此相随，再如图 10 所示，还存储用作被检索关键词的在各短语中所包含的加重字句数。此外，作为韵律信息存储的停顿信息不限于短语前后的停顿时间，也可以包括加重字句前后的停顿时间。

基频图案检索部 141、语音强度图案检索部 44L、音韵持续时间图案检索部 443 以及停顿信息检索部 444，都是以短语单位进行韵律信息的检索，因而作为近似成本也可以考虑短语中所含加重字句。再有，除停顿信息检索部 444 之外，与检索出的基频图案等以及近似成本在一起还输出检索键与被检索键的音素序列中各音素的一致度。而另一方面，停顿信息检索部 444 则连同停顿信息与近似成本还输出各加重字句的拍数与重音位置等。

基频图案变形部 451、语音强度图案变形部 452 与音韵持续时间图案变形部 453 与实施形式 1 ~ 3 的韵律信息变形部 150 等相同，应用韵律信息变形规则存储部 460 中所存储的规则，根据基频图案检索部 441 等输出的近似成本进行韵律信息的变形，同时还根据检索关键词与被检索关键词的音素序列中各音素的一致度进行变形。具体地说，例如相对于日语“たかな”而用到了“さかな”这类只有部分音素不同的语词的韵律信息时，易使不同音素的语音强度图案如图 2 中符号 P 所示部分那样减弱，而变形成为使音素的不同难以产生显著影响。此外，也可不必要有这种根据各音素的一致度进行的变形，或也可以不进行根据近似成本的变形而只进行根据各音素一致度的变形。

停顿时间变更部 454 应用韵律信息变形规则存储部 460 所存储的规则，根据停顿信息检索部 444 输出的近似成本进行韵律信息的变形，同时再根据各加重字句的拍数和重音位置等一致度来变更停顿时间。

如上所述，通过以短语为单位进行韵律信息的检索或变形等，就能发出遵从文章顺序的较自然的合成语音。此外，与实施形式 2 相同，通过采用由韵律信息数据库 430 检索出的停顿信息，可以发出停顿时间较自然的合成语音，同时与实施形式 3 相同，通过采用另外的近似成本独立地进行基频图案等的检索与变形，能够分别根据各最适当的基频图样进行语音合成，而减小韵律信息数据库 430 的存储容量。此外，通过进行根据各音素一致度的基频图案等的变形就易使音素不同的影响显著减弱，再根据各加重字句的拍数或重音位置等一致度来变更停顿时间，这样便可发出停顿时间更自然的合成语音。

#### (实施形式 5)

下面以把音韵类别序列用于韵律信息的检索中为例说明实施形式 5 的语音合成系统。

图 11 为示明实施形式 5 的语音合成系统结构的功能框图。图 12 为示明音韵分类例子的说明图。

上述音韵类别是把音韵通过据各音韵间的语音学的特征求得的距离，即由各音韵的调音方式、调音位置、持续时间长短等，加以分组的结果。具体地说，由于使这种音韵类别为同一类别的音素相互之间具有相似的音响特性，例如某个加重字句和将其中一部分音素代换为相同音韵类别的其他音素所得的加重字句，多数具有相同或较相似的韵律信息。于是，即使在韵律信息的检索中的音素序列不一致时，在各音素的音韵类别不一致的情形，纵然乱用了韵律信息，许多场合下也能发出适当的合成语音。再有，音韵的分组化也不限于以上所述，例如图 12 所示，可以依照应用据音韵相互间的异听表的多变量分析等所确定的音韵间的距离（心理距离）将音韵分组，或根据音韵的物理特性（音韵的基频、强度、持续时间以及音频频谱等）的相似度分组，或是应用多变量分析等的统计方法将韵律图案分组，但为了最好地反映上述韵律图案的组别，也可用统计方法将音韵分组。

具体说明如下。此实施形式 5 的语音合成系统与实施形式 1 的语音合成系统相比，代替韵律信息数据库 130 而配备了韵律信息数据库 730，另外不同的是设置了音韵类别序列生成部 790。

在上述韵律信息数据库 730 中，除实施形式 1 的韵律信息数据库 130 的存储内容之外，还存储有示明了加重文节的各音素所属音韵类别的音韵类别序列用作被检索关键词。这里用作音韵类别序列的具体表示形式，例如可以表示为分配给各音韵类别的序号或符号序列，或是把各音韵类别内的某一音素作为代表音素，或是表示为这种代表音素的序列。

音韵类别序列生成部 790 将语言处理部 120 输出的各加重字句的表音符号序列变换为音韵类别序列输出。

韵律信息检索部 740 根据音韵类别序列生成部 790 输出的音韵类别序列以及语言处理部 120 输出的各加重字句的表音符号序列与语言信息，检索韵律信息数据库 730 的韵律信息，以输出检索出的韵律信息与近似成本。上述近似成本由于包含有音韵类别序列的一致程度（例如各音韵的音韵类别的相似程度），因而即使例如在音素序列不一致时，由于音韵类别序列一致情形下能取小的值，从而能检索（选择）更适当的韵律信息，发出自然的合成语音。此外，通过首先将检索候补集中于音韵类别序列一致的或相似的集合，便容易提高检索速度。

此外，上述例子中所示的是将语音处理部 120 输出的表音符号序列通过音韵类别序列生成部变换为音韵类别序列，但并不局限于此，而是可以使语言处理部 120 具有生成音韵类别序列的功能，或是使韵律信息检索部 740 具有将输入的表音符号序列变换为音韵类别序列的功能，等等。若是韵律信息检索部 740 具有将从韵律信息数据库读出的音素序列变换为音韵类别序列的功能，则也可以采用与实施形式 1 的韵律信息数据库 130 相同的未存储音韵类别序列的韵律信息数据库。

也可不限于将音素序列与音韵类别序列共同用作检索关键词而

只采用音韵类别序列，此时由于能只汇集音素序列不同的韵律信息，就可减小数据库容量和提高检索速度等。

上述各实施形式与变形例等说明的结构部件也可取种种组合形式。具体地说，将实施形式 5 所示的音韵类别序列用于韵律信息的检索等之中的方法，也适用于其他的实施形式等。

实施形式 3、4 所示的与对应于各音素一致度的韵律信息变形，在其他的实施形式等之中也可代之以对应于近似成本的变形，或与它同时使用。此外还可采用各音素、各拍、各音节、波形生成部中各语音波形的生成单位、各音韵的一致度等进行变形。也可选择根据变形的韵律信息所用的一致度。具体地说，例如在基频图案的变形中，可应用近似成本或各音素等的一致度中的某一个，而在音素强度图案的变形中则可同时用到这两方。上述音素等的一致度，例如可基于由基频与强度、持续时间、基于音频频谱等特性的距离、调音方式、调音位置、持续时间等于语音学意义下求得的距离，或基于根据所取实验的异听表所得的距离等，加以确定。

实施形式 5 所示的音韵分类用于检索等之中的方法，在其他实施形式等之中也可用音素序列代替，或也可与之同时使用。

如实施形式 2、4 所示，将停顿信息作为韵律信息存储于韵律信息数据库以进行检索的结构，也适用于其他实施形式等，相反，在实施形式 2、4 等之中，停顿信息也可用于检索。

语言处理部并非是必要的，表音符号序列等也可直接从外部输入。这种结构，例如特别适用于携带式电话之类的小型设备，使得设备的小型化或通信数据的压缩等更容易。还可从外部输入表音符号序列和语言信息。具体地说，例如采用大型服务器进行精度高的语音处理，将其结果输入时就能发出更适当的语音。另一方面，若只简单地采用表音符号序列等，则可简化结构。

用于合成语音的韵律信息不限于以上所述情形。例如，取代音韵持续时间等也可采用音素持续时间图案、拍持续时间图案、音节持续时间图案等。此外，也可采用包含上述这类持续时间在内的种

种韵律信息的组合结果。

韵律控制单位即韵律信息的存储、检索、变形等单位，既可以用加重字句或1个以上字句组成的短语，也可以用文节、单词、加重文节单位或1个以上文节、单词、加重字句单位组成的短语单位，或也可以上述混合组成的结果。此外，与韵律控制单位（例如1个以上加重语句组成的短语）不同，也可以将其他各单位（例如加重字句）的拍数与重音位置等的一致度用于韵律信息的变形等之中。

检索关键词的项目与数量也不限于以上所述。即一般地说，检索关键词项目多时容易检索出合适的候选，但是为了容易地检索出最适当的候选，则最好要各项目的一致度的决定与加权的各式等都能最优化。此外，为了简化结构和提高处理速度，也可略去与检索精度关联性小的检索键。

在以上各例中虽是以日语为例进行说明，但本发明并不局限于此而能同样容易地用于种种不同的语言。这时可以施加相应于各种语言特性的变形，例如可对拍单位的处理施加拍或音节单位的处理等的变形。又，在韵律信息数据库130等之中也可存储有关多种语言的信息。

此外，上述结构既可以由计算机（以及外围设备）与程序作实际组装，也可以通过硬件实际安装。

如上所述，根据本发明，例如可将从实际语音提取的基频图案与语音强度图案、语音持续时间图案、停顿信息等的韵律信息存储于数据库中，相对于文本或表音符号序列等输入的发音目标，例如从数据库中检索使近似成本为最小的韵律信息，根据近似成本与一致度等，通过根据所定的变形规则使选择出的韵律信息变形，发出对应于任意输入文本等的自然的合成语音。特别是当语音信息数据库内即使不存在对应于输入文本等的发音内容时，也能有效地以同样的音质发出，即整体上看能发出接近实际语音的自然合成语音。

于是，本发明能用于家电制品以及汽车驾驶导向系统、手机等种种电子设备中，对设备状态与操作等指示、应答消息等信息进行

---

语音表示，而在个人计算机等之中则能用于语音接口进行的操作、光学字符识别（OCR）的字符识别结果的确认等，即本发明可以有效地应用于上述这类领域中。

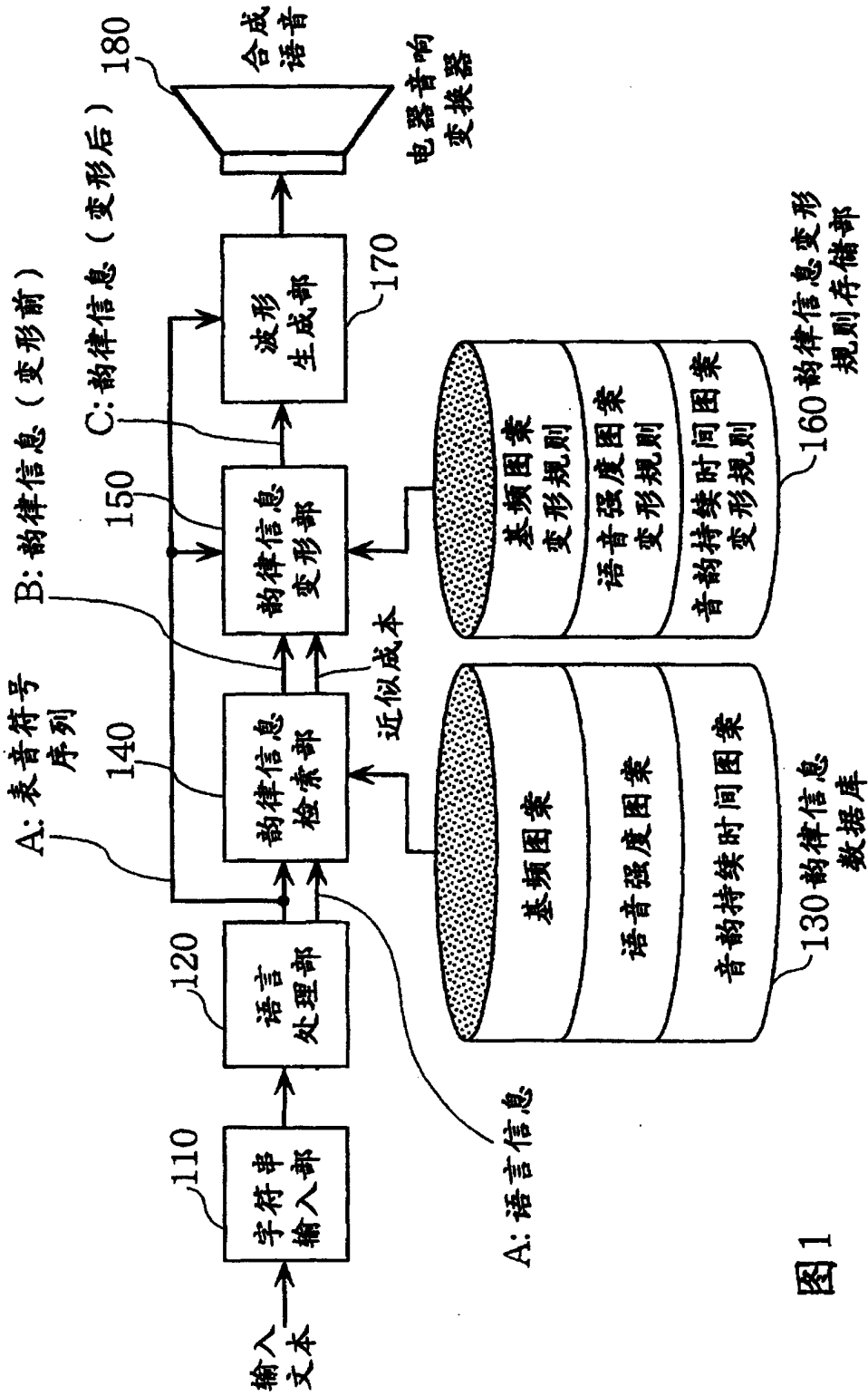


图 1

各加重字句 的输入字符串	A: 语言处理部 (120) 的输出		B: 韵律信息检索部 (140) 的输出		C: 韵律信息变形部 (150) 的输出	
門真市	表音 符号序列	kadoma'shi/0	基频图案		基频图案	
	语言信息 (语法信息 语义信息)	名词 地名 (住所)	语音强度 图案		语音强度 图案	
門真	表音 符号序列	ka'doma/50	音韵持续 图案		音韵持续 图案	
	语言 信息	名词 地名 (住所)	基频图案		基频图案	
6番地までの	表音 符号序列	rockkuba' Nclchimadeno/0	语音强度 图案		语音强度 图案	
	语言 信息	名词/副助词/ 格助词/地名 (住所)/终点	音韵持续 图案		音韵持续 图案	

图 2

韵律信息数据库 (130) 的存储内容

被检索关键词				韵律信息						
音素序列	重音位置	拍数	紧前面的停顿时间	紧后面的停顿时间	语法信息·意义信息 词类别	音节信息	基频图案	语音强度图案	音韵持续时间	持续图案
minamino kaijo-niwa	9	10	文首	短 ( $>0$ ) ( $\leq 100$ )	名词/ 格助词/ 名词/ 格助词/ 副助词				t:87 n:82 a:78 i:79 n:65 o:66 k:64 a:64 j:62 o:63 n:65 w:67 a:52	
te-kiatsuqa	3	6	短 ( $>0$ ) ( $\leq 100$ )	无 (0)	名词/ 格助词	连用形; 格要素; 天气			t:66 e:67 -:67 c:26 k:60 a:67 c:15 ts:18 u:18 c:49	
ari	1	2	无 (0)	长 ( $>100$ )	动词	连用形中 止; 存在			a:72 i:76	
hoNshu-	1	4	长 ( $>100$ )	无 (0)	名词	连体形1/2 ; 地名				

图 3

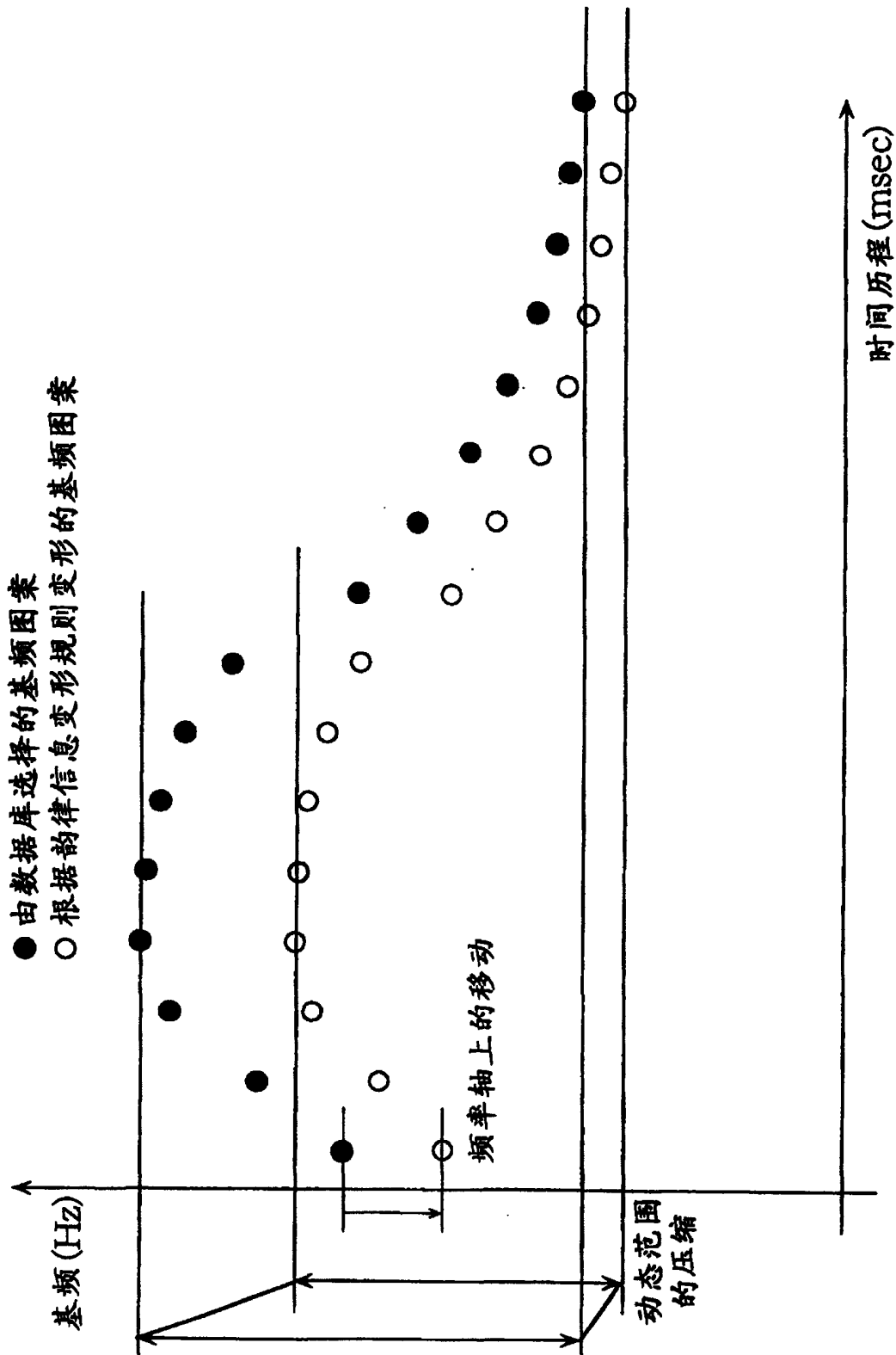


图 4

输入文本“門真市門真6番地までの所要時間は35分です。”

输入文本	检索关键词的表序列 符号序列	被检索词的表序列 符号序列	根据格检索项目的一致程度所希望的 韵律信息的变形方式(值的增减)						综合的 所希望的 变形方法				
			音素序列	重音	拍数	紧前的 停顿时间	紧后的 停顿时间	词类别	文 信息				
門真市	kadoma'shi	nagoya'shi	—	—	—	—	—	—	—	—	—	—	—
3	sa'Nju-	ka'Nclto-	—	—	—	↘	—	—	—	—	↗	↗	↗
5分です	go'fuNdesxu	na'ruNdesxu	—	—	—	—	—	—	—	↘	↗	↗	↗

图5

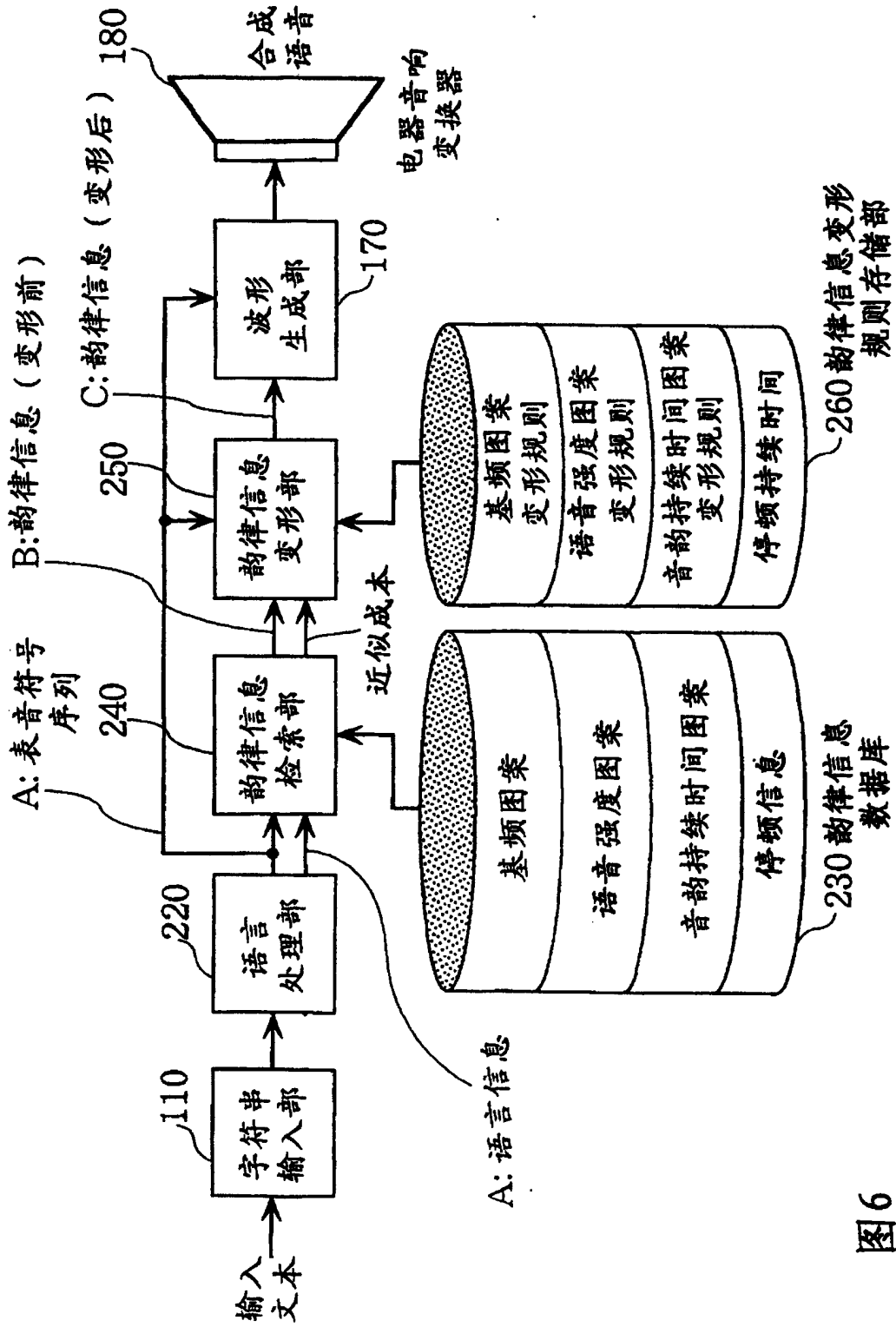


图6

韵律信息数据库 (230) 的存储内容

		被检索关键				韵律信息			
音素序列	重音位置	拍数	语法信息·意义信息		基频图案	语音强度图案	音韵持续时间	停顿信息	
			词类	文节信息				紧前的停顿时间	紧后的停顿时间
minamino kaijo-niwa	9	10	名词/ 格助词/ 名词/ 格助词/ 副助词	连体形; 方向/连 用形; 地点			t:187 n:183 a:178 i:179 n:165 o:155 k:122 a:124 j:162 o:89 f:87 w:87 a:52	34	34
te-kiatsuqa	3	6	名词/ 格助词	连用形 格要素; 天气			t:17 e:69 -c:76 k:53 i:60 a:82 ts:18 u:58 g:47 a:49	34	0
ari	1	2	动词	连用形终 止; 存在			a:72 i:76	0	122
hoNshu-	1	4	名词	连体形1/ 2; 地名			h:26 o:53 N:72 sh:36	122	0

图7

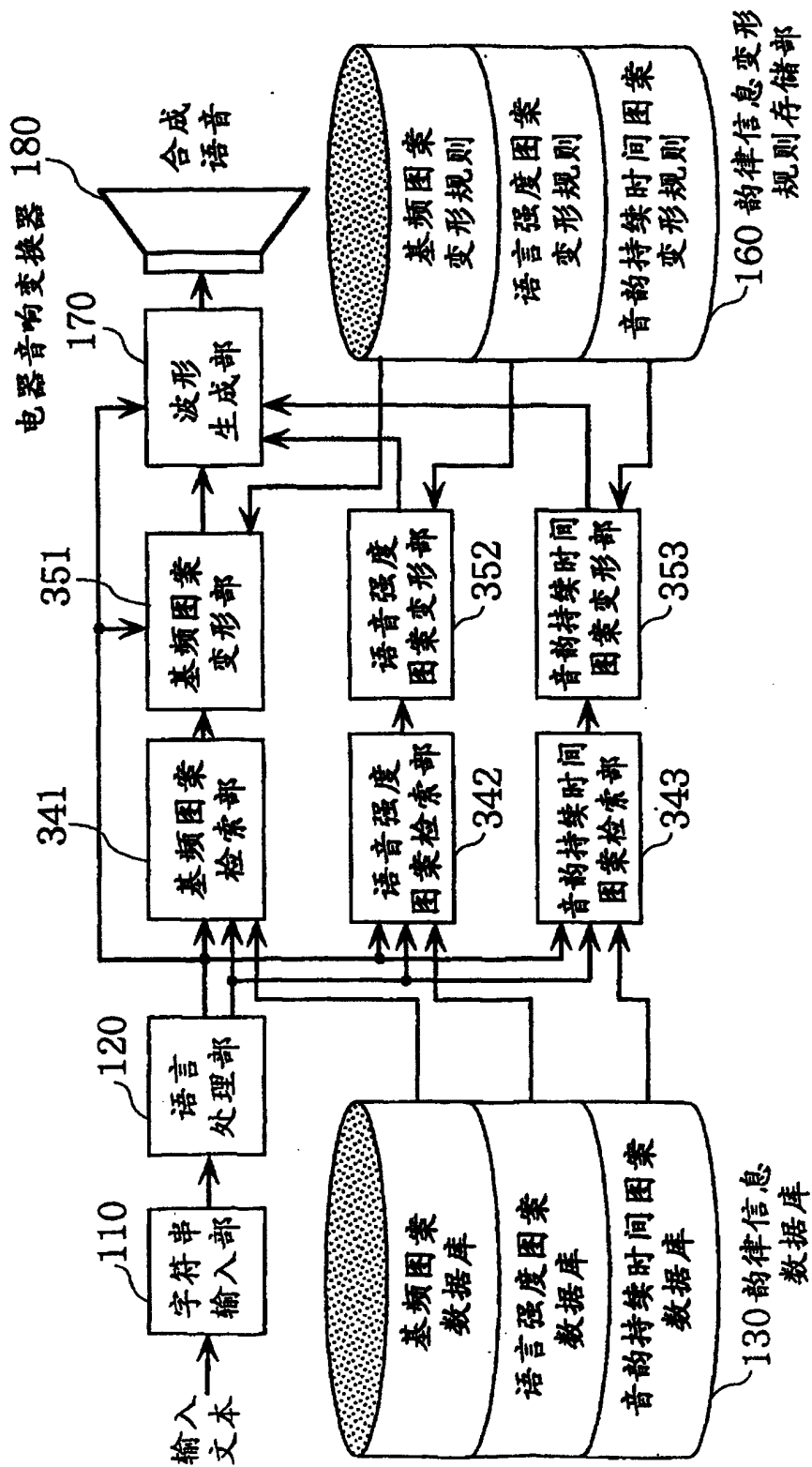


图8

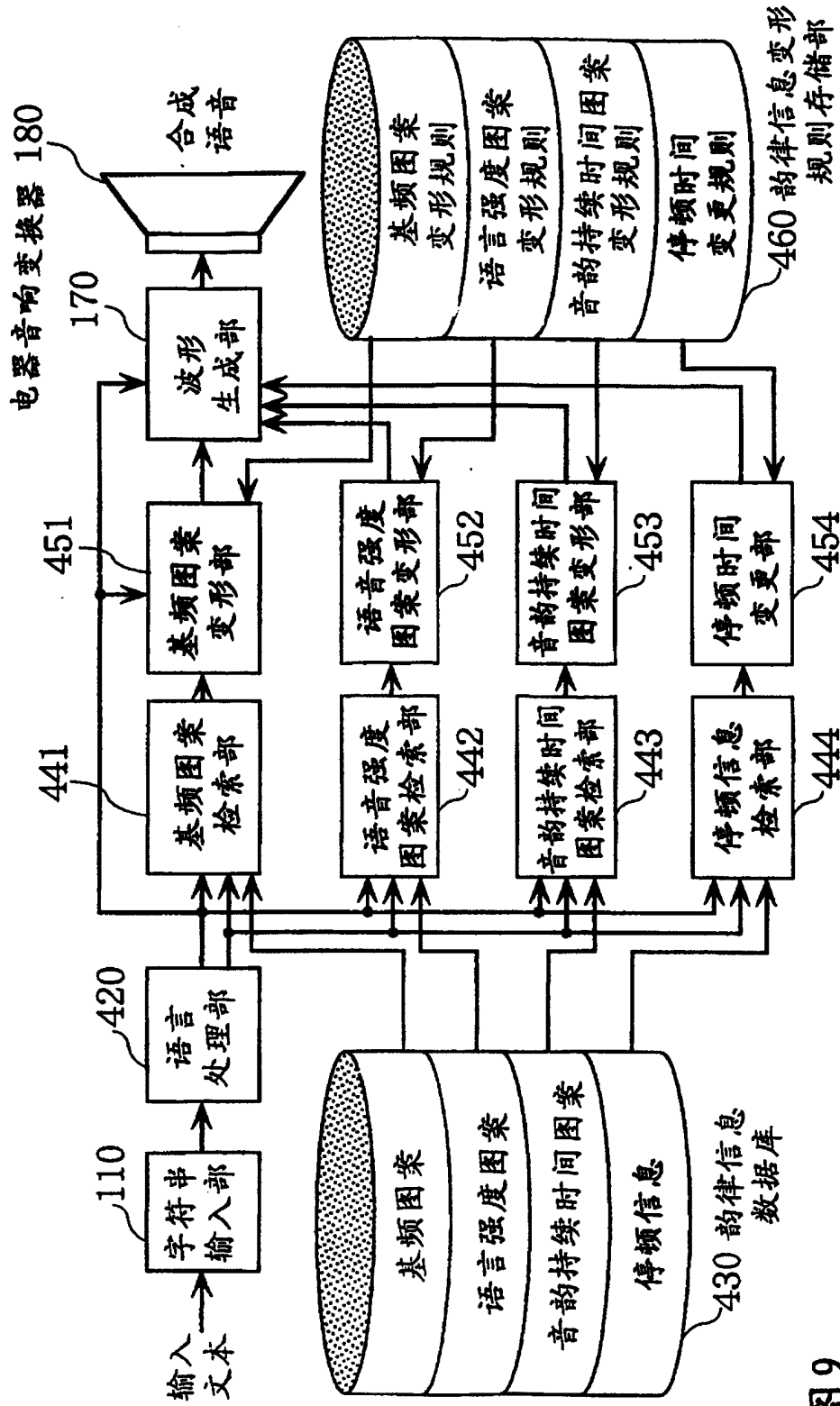


图9

音律信息数据库(130)的存储内容

被检索关键词						音律信息	
音素序列	重音位置	加重句数	拍数	语法信息·意义信息		紧前的停顿时间	紧后的停顿时间
				词类	文节信息		
minamino kaijo-niwa	9	1	10	名词/ 格助词/ 名词/ 格助词/ 副助词	连体形; 方向/连 用形; 地点	文首	34
te-kiatsuga /ari	3,7	2	8	名词/ 格助词 动词	连用形格 要素; 天 气/适用 形中止; 存在	34	122
hoNshu-/ clchu-buno/ cltaihe-yo- gawawa	1,5,10,11	3	17	名词/名词 格助词/ 名词/ 格助词/ 副助词	连体形; 地名地点/ 适用形; 地名地点	122	0

图10



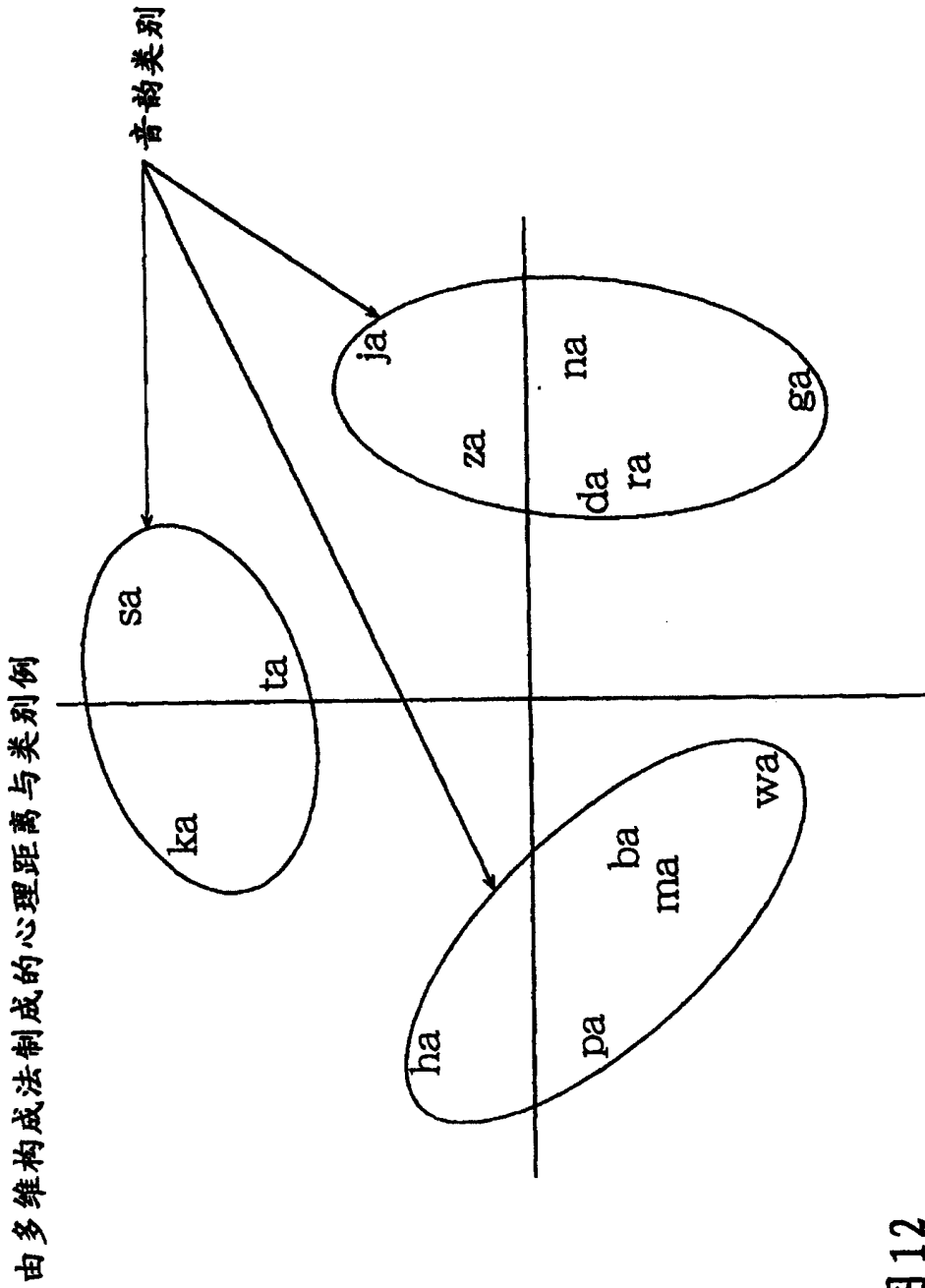


图12

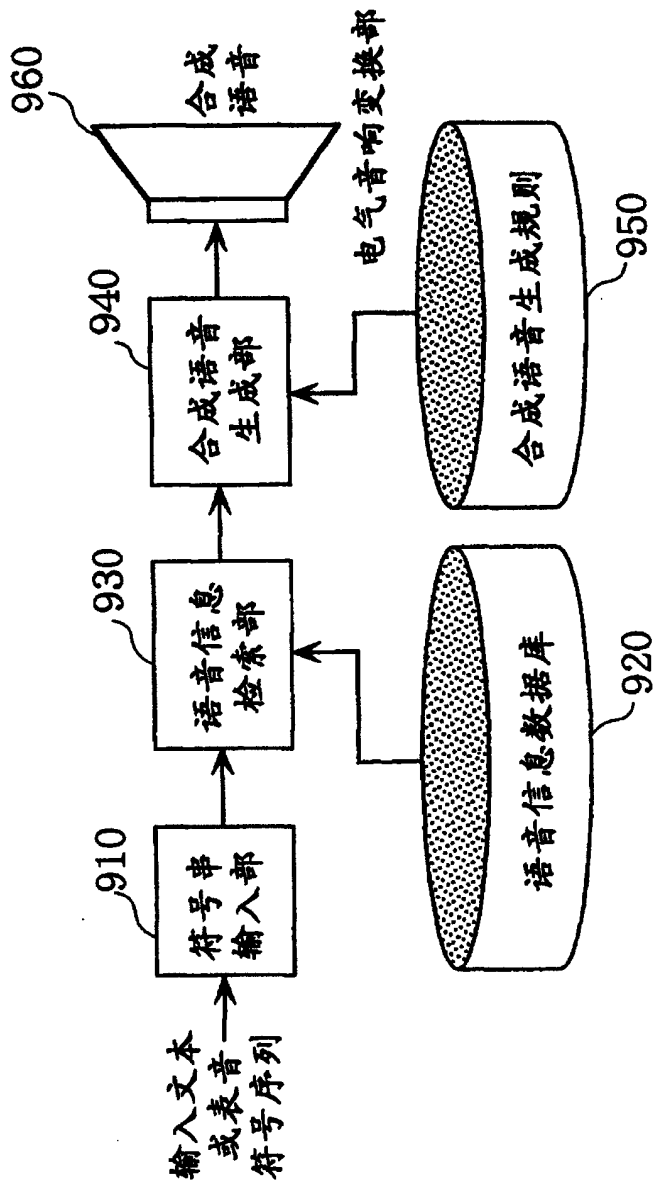


图13