

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G01L 13/04 (2006.01)

G10L 13/08 (2006.01)



[12] 发明专利说明书

专利号 ZL 200510007454.2

[45] 授权公告日 2007 年 9 月 12 日

[11] 授权公告号 CN 100337104C

[22] 申请日 2005.2.21

[21] 申请号 200510007454.2

[30] 优先权

[32] 2004. 2. 20 [33] JP [31] 2004 - 044852

[73] 专利权人 雅马哈株式会社

地址 日本静岡県

[72] 发明人 川原毅彦 剑持秀纪

[56] 参考文献

CN 1297561A 2001.5.30

JP 2000 - 339137A 2000.12.8

CN 1208910A 1999.2.24

CN 1136198A 1996.11.20

CN 1378199A 2002.11.6

JP 8 - 146965A 1996.6.7

审查员 刘红梅

[74] 专利代理机构 北京天昊联合知识产权代理有限公司

代理人 何立波 张天舒

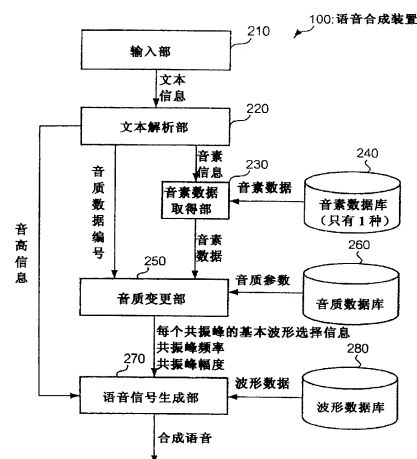
权利要求书 3 页 说明书 11 页 附图 12 页

[54] 发明名称

语音合成装置、方法和记录语音合成程序的
记录介质

[57] 摘要

本发明提供一种即使在对硬件资源有较大限制的环境下,也能够生成各种音质的合成语音的语音合成装置。在具有一种音素数据的语音合成装置(100)中,设置音质变更部(250)和音质数据库(260)。音质变更部(250)将从文本解析部(220)提供的音质数据编号作为检索关键字,对音质数据库(260)进行检索,取得音质参数。音质变更部(250)根据所取得的音质参数,对由音素数据取得部(230)取得的音素数据所表示的各音素的音质进行变更。



1. 一种语音合成装置，其特征在于，具有：

获取单元，从被输入语音合成装置的文本信息，获取用于指定合成语音的音素的音素指定信息和用于指定该合成语音的音质的音质指定信息；

第一存储单元，用于存储表示各音素的多个音素数据，这各个音素数据包含由成对的共振峰频率和共振峰幅度构成的共振峰信息；

第二存储单元，用于存储多个音素数据加工信息，这些音素数据加工信息是用于变更上述各音素的音质的信息，这各个音素数据加工信息包含表示共振峰频率的变更量的信息和表示共振峰幅度的变更量的信息；

第一抽取单元，用于从上述第一存储单元抽取出与上述音素指定信息所表示的音素对应的音素数据；

第二抽取单元，用于从上述第二存储单元抽取出与上述音质指定信息所表示的音质对应的音素数据加工信息；以及

生成单元，用于根据上述抽取出的音素数据加工信息中包含的表示共振峰的变更量的信息，变更上述抽取出的音素数据中包含的共振峰频率，且根据上述抽取出的音素数据加工信息中包含的表示共振峰幅度的变更量的信息，变更上述抽取出的音素数据的共振峰幅度，由此，对上述抽取出的音素数据进行加工，根据上述加工后的音素数据生成上述合成语音。

2. 根据权利要求1所述的语音合成装置，其特征在于，

上述生成单元根据上述加工后的音素数据，生成每个共振峰的共振峰波形，并通过将生成的各共振峰波形相加，生成上述合成语音。

3. 根据权利要求2所述的语音合成装置，其特征在于，

上述获取单元从上述文本信息中除了获取上述音素指定信息和上述音质指定信息之外，还获取用于指定上述合成语音的音高的音高

指定信息，

上述生成单元还向将生成的各共振峰波形相加而成合成共振峰波形，赋予上述音高指定信息所表示的音高，由此生成上述合成语音。

4. 根据权利要求1所述的语音合成装置，其特征在于，上述获取单元从上述文本信息中抽取出关键字，然后根据所抽取出的关键字，来判定适合于上述文本信息的音质。

5. 根据权利要求1所述的语音合成装置，其特征在于，
该各音素数据加工信息，还包括表示共振峰频率的振荡速度的信息及表示共振峰频率的振荡幅度的信息，

前述生成单元，根据上述表示共振峰的振荡速度的信息及上述表示共振峰频率的振荡幅度的信息，将变更后的共振峰振荡。

6. 一种语音合成方法，其特征在于，具有以下步骤：

获取步骤，从被输入语音合成装置的文本信息，获取用于指定合成语音的音素的音素指定信息和用于指定该合成语音的音质的音质指定信息；

第一抽取步骤，从第一存储单元抽取出与上述音素指定信息所表示的音素对应的音素数据，该第一存储单元用于存储表示各个音素的多个音素数据，这各音素数据包含由成对的共振峰频率和共振峰幅度构成的共振峰信息；

第二抽取步骤，用于从第二存储单元抽取出与上述音质指定信息所表示的音质对应的音素数据加工信息，该第二存储单元用于存储多个音素数据加工信息，这些音素数据加工信息是用于变更上述各音素的音质的信息，这各个音素数据加工信息包含表示共振峰频率的变更量的信息和表示共振峰幅度的变更量的信息；以及

生成步骤，用于根据上述抽取出的音素数据加工信息中包含的表示共振峰的变更量的信息，变更将上述抽取出的音素数据中包含的共振峰频率，且根据上述抽取出的音素数据加工信息中包含的表示共

振峰幅度的变更量的信息，变更上述抽取出的音素数据的共振峰幅度，由此，对上述抽取出的音素数据进行加工，根据上述加工后的音素数据生成上述合成语音。

语音合成装置、方法和记录语音合成程序的记录介质

技术领域

本发明涉及一种根据所输入的文本信息而生成合成语音的语音合成装置、语音合成方法和记录语音合成程序的记录介质。

背景技术

图 17 是表示根据所输入的文本信息生成合成语音的现有语音合成装置 100 结构的图。

输入部 110 在从未图示的操作部等接收了「こんにちは」等文本信息之后，将该文本信息提供给文本解析部 120。文本解析部 120 利用单词辞典等，对所接收到的文本信息进行单词解析、语法解析等，生成表示「こ」、「ん」、「に」、「ち」、「わ」这些短音节单位的各音素的音素信息，以及表示各音素长度、高度、强度的韵律信息，然后提供给语音合成部 130。语音合成部 130 根据从文本解析部 120 提供的各音素信息，从音素数据库 140 中取得短音节单位的语音数据（以下称为音素数据）。然后，语音合成部 130 按照韵律信息，对所取得的各音素数据进行适当加工・连接等，生成合成语音信号，然后作为合成语音而从扬声器等输出。用户通过聆听从语音合成装置输出的合成语音，可以确认所输入的文本信息的内容。

但是，在上述音素数据库中，仅登录了特定朗读者（例如男性朗读者）发出的 1 种音素数据。因此，例如在利用具有上述特定朗读者音质的合成语音输出年轻女性等喜欢使用的文本信息（「ちょう…」或「…みたいなあ」等文本信息）的情况下，会出现用户感觉音质和语音内容之间有差异感的问题。

为了解决上述问题，这样的技术被提出，即在音素数据库中预先登录多种音素数据（例如男性、女性、儿童、老人等不同的音素数据），按照所输入的文本信息的内容等来选择最适宜的音素数据，然

后利用所选择的音素数据来生成合成语音（例如，参照专利文献 1）。

专利文献 1：特开 2000-339137 号公报（第 3~4 页）

发明内容

根据专利文献 1 所公开的技术，确实可以获得合适的合成语音，但为了实现合成语音，必须在音素数据库中登录多种音素数据。但是会出现这样的问题，即在对存储器和 CPU 等硬件资源有较大限制的移动终端等中无法安装这样的多种音素数据，其结果是移动终端等无法生成各种音质的合成语音。

本发明就是鉴于上述问题而提出的，其目的在于，提供一种即使在对硬件资源有较大限制的环境下，也能生成各种音质的合成语音的语音合成装置、语音合成方法和记录语音合成程序的记录介质。

为了解决上述问题，本发明的语音合成装置，其特征在于，具有：获取单元，从被输入语音合成装置的文本信息，获取用于指定合成语音的音素的音素指定信息和用于指定该合成语音的音质的音质指定信息；第一存储单元，用于存储表示各音素的多个音素数据；第二存储单元，用于存储多种音素数据加工信息，这些音素数据加工信息是用于变更上述各音素的音质的信息，它们表示音素数据的加工内容；第一抽取单元，用于从上述第一存储单元抽取出与上述音素指定信息所表示的音素对应的音素数据；第二抽取单元，用于从上述第二存储单元抽取出与上述音质指定信息所表示的音质对应的音素数据加工信息；以及生成单元，用于根据上述抽取出的音素数据加工信息，对上述抽取出的音素数据进行加工，生成上述合成语音。

本发明的语音合成装置，优选上述各音素数据包含表示音素的共振峰的共振峰信息，上述音素数据加工信息包含表示上述共振峰的变更内容的共振峰变更信息，上述生成单元根据上述共振峰变更信息，对上述共振峰信息进行变更，然后根据对变更后的各共振峰进行加法运算后的信号波形，生成上述合成语音。

本发明的语音合成装置，优选上述共振峰信息由成对的共振峰频率和共振峰幅度构成，在上述共振峰变更信息中，包含有用于表示

上述共振峰频率的变更内容的共振峰频率变更信息和用于表示上述共振峰幅度的变更内容的共振峰幅度变更信息,上述生成单元根据上述共振峰频率变更信息和上述共振峰幅度变更信息,分别对上述音素数据所表示的音素的各共振峰频率和各共振峰幅度进行变更,由此得到上述变更后的各共振峰。

本发明的语音合成装置,优选上述获取单元从上述文本信息中除了获取上述音素指定信息和上述音质指定信息之外,还获取用于指定上述合成语音的音高的音高指定信息,上述生成单元向对上述变更后的各共振峰进行加法运算之后的信号波形赋予上述音高指定信息所表示的音高,由此获得上述合成语音。

本发明的语音合成装置,优选上述文本信息包含上述音质指定信息,上述获取单元从上述文本信息中获取上述音质指定信息。

本发明的语音合成装置,优选上述获取单元从上述文本信息中抽取出关键字,然后根据所抽取出的关键字,来判定适合于上述文本信息的音质。

本发明还提供一种语音合成方法,其特征在于,具有以下步骤:获取步骤,从被输入语音合成装置的文本信息,获取用于指定合成语音的音素的音素指定信息和用于指定该合成语音的音质的音质指定信息;第一抽取步骤,从第一存储单元抽取与上述音素指定信息所表示的音素对应的音素数据,该第一存储单元用于存储表示各音素的多个音素数据;第二抽取步骤,用于从第二存储单元抽取与上述音质指定信息所表示的音质对应的音素数据加工信息,该第二存储单元用于存储多种音素数据加工信息,这些音素数据加工信息是用于变更上述各音素的音质的信息,它们表示音素数据的加工内容;以及生成步骤,用于根据上述抽取出的音素数据加工信息,对上述抽取出的音素数据进行加工,生成上述合成语音。

发明的效果

如上所述,根据本发明,即使在对硬件资源有较大限制的环境下,也能够生成各种音质的合成语音。

附图说明

图 1 是表示本实施方式的语音合成装置的功能性结构的框图。

图 2 是例示出该实施方式的文本信息的图。

图 3 是例示出该实施方式的音素数据库的登录内容的图。

图 4 是例示出该实施方式的音素数据结构的图。

图 5 是用于说明包含在该实施方式的音素数据中的各帧信息的图。

图 6 是例示出该实施方式的音质数据库登录内容的图。

图 7 是表示该实施方式的音质参数的一个结构例的图。

图 8 是表示该实施方式的音质变更处理的流程图。

图 9 是例示出该实施方式的映射函数的图。

图 10 是表示该实施方式的男性的音素分析结果的图。

图 11 是表示该实施方式的的女性音素的分析结果的图。

图 12 是例示出该实施方式的振荡表的图。

图 13 是例示出从该实施方式的振荡表中读出的振荡值与时间的关系的图。

图 14 是用于说明该实施方式的共振峰频率振荡的图。

图 15 是用于说明该实施方式的音高赋予处理的图。

图 16 是例示出该实施方式的被进行了音质变更处理和音高赋予处理后的特定共振峰的共振峰波形的图。

图 17 是表示现有的语音合成装置的功能性结构的图。

具体实施方式

以下参照附图，对本发明的实施方式进行说明。

A. 本实施方式

图 1 是表示本实施方式的语音合成装置 100 的功能性结构的图。在本实施方式中，假定语音合成装置 100 安装在移动电话或 PHS (Personal Handyphone System)、PDA (Personal Digital Assistance) 等对硬件资源有较大限制的移动终端中的情况，但本发

明不限于此，可以应用在各种电子设备中。

输入部 210 将经由未图示的操作部等输入的文本信息提供给文本解析部 220。图 2 是例示了文本信息的图。

文本内容信息是用于表示应作为合成语音输出的文本的内容（例如「こんにちは」）的信息。另外，在图 2 中示出了仅由平假名表示的文本内容信息，但文本内容信息不限于平假名，也可以是由汉字、罗马字、片假名等各种文字和各种记号表示的信息。

音质数据编号（音质指定信息）是用于分别识别后述的多个音质参数（音素数据加工信息）的唯一编号（图 2 中的 K1~Kn）。在本实施方式中，通过适当选择・利用该音质参数，可以根据特定朗读者（在本实施方式中假定为「男性朗读者」）发出的一种音素数据，获得各种音质的合成语音（后面详细说明）。

音高（pitch）信息（音高指定信息）是用于给合成语音赋予音高（换句话说，指定合成语音的音高）的信息，它由指定「C（哆）」～「B（唏）」等音阶的信息构成（参照图 2）。

文本解析部 220 对从输入部 210 提供的文本信息进行解析，然后将解析结果分别提供给音素数据取得部 230、音质变更部 250、语音信号生成部 270。具体地讲，在被提供了图 2 所示的文本信息之后，文本解析部 220 首先将「こんにちは」这样的文本内容信息分解为「こ」、「ん」、「に」、「ち」、「わ」这样的短音节单位的音素。所谓的短音节，是指表示读音单位，基本由 1 个辅音和 1 个元音构成的音节。

文本解析部（获取单元）220 在按照这样的方式将文本内容信息分解为短音节单位的音素之后，生成用于指定这些合成语音的各音素的音素信息（音素指定信息），然后依次提供给音素数据取得部 230。接下来，文本解析部 220 从文本信息中分别取得音质数据编号（例如 K3）、音高信息（例如 C（哆）），然后将所取得的音质数据编号提供给音质变更部 250，而将所取得的音高信息提供给语音信号生成部 270。

音素数据取得部（第一抽取单元）230 将从文本解析部 220 提供

的音素信息作为关键字，对音素数据库 240 进行检索，由此取得与音素信息所表示的音素对应的音素数据。图 3 是例示出音素数据库 240 的登录内容的图。如图 3 所示，在音素数据库（第一存储单元）240 中，除了登录有表示一个男性朗读者的短音节单位的各音素（「あ」、「い」、…「ん」等）的一系列音素数据 1~m 之外，还登录有该一系列音素数据的数量（以下称为登录音素数据数量）等。

图 4 是例示出表示某一音素（例如「こ」等）的音素数据的结构图，图 5 是用于说明包含在音素数据中的各帧信息的图。另外，图 5 的 A 示出了上述男性朗读者朗读某一音素（例如「こ」等）时的语音波形 vw 与各帧 FR 之间的关系，图 5 的 B、图 5 的 C 和图 5 的 D 分别示出了对于第 1 帧 FR1、第 2 帧 FR2、第 n 帧 FRn 的共振峰分析结果。

如图 4 所示，音素数据由第 1 帧信息~第 n 帧信息构成。各帧信息具有对对应的各帧 Fr（参照图 5）进行共振峰分析而得到的第 1 共振峰信息~第 k 共振峰信息，以及用于表示各帧 FR 的语音是浊音（voiced sound）还是清音（voiceless sound）的浊音/清音判别标记（例如，“1”=浊音，“0”=清音）。

构成各帧信息的第 1 帧信息~第 k 帧信息由表示对应共振峰的、成对的共振峰频率 F 和共振峰幅度 A 构成（参照图 5 的 B~图 5 的 D）。例如，构成第 1 帧信息的第 1 共振峰信息~第 k 共振峰信息分别由（F11, A11）、（F12, A12）、…（F1k, A1k）这些成对的共振峰频率和共振峰幅度构成（参照图 5 的 B），……，构成第 n 帧信息的第 1 共振峰信息~第 k 共振峰信息分别由（Fn1, An1）、（Fn2, An2）、…（Fnk, Ank）这些成对的共振峰频率和共振峰幅度构成（参照图 5 的 D）。

音素数据取得部 230 在根据从文本解析部 220 提供的各音素信息（表示「こ」、「ん」、「に」、「ち」、「わ」等的各音素信息）取得对应的各音素数据之后，将这些音素数据提供给音质变更部 250。

音质变更部 250 变更由音素数据取得部 230 取得的各音素数据所表示的音素的音质。具体地讲，音质变更部（第二抽取单元）250

首先将文本解析部 220 提供的音质数据编号作为检索关键字,对音质数据库(第二存储单元)260 进行检索,取得对应的音质参数。然后,音质变更部 250 根据所取得的音质参数,进行上述各音素的音质的变更。

图 6 是例示出音质数据库 260 的登录内容的图。

如图 6 所示,在音质数据库(第二存储单元)260 中,作为用于变更上述各音素的音质的必要信息,存储有表示音素数据的加工内容的多种音质参数 1~L,以及表示该音质参数的数量的登录数量信息。

图 7 是表示音质参数的一个结构例的图。

如图 7 所示,音质参数(音素数据加工信息)具有用于确定该参数的音质数据编号、以及表示是否变更合成语音的性别的性别变更标记、表示第 1~第 k 共振峰的变更内容的第 1~第 k 共振峰变更信息。其中,在例如将上述性别变更标记设定为“1”的情况下,由音质变更部 250 进行用于变更合成语音的性别的处理(以下称为性别变更处理),而在将上述性别变更标记设定为“0”的情况下,不进行上述性别变更处理(后面详细说明)。另外,在本实施方式中,由于假定男性朗读者发出的 1 种音素数据,所以在将该性别变更标记设定为“1”的情况下,合成语音的特征从男性特征变更为女性特征。另一方面,在性别变更标记被设定为“0”的情况下,合成语音的特征还保持男性的特征而不变更。

另一方面,各共振峰变更信息具有用于选择后述的各共振峰基本波形(正弦波等)的基本波形选择信息、表示各共振峰频率的变更内容的共振峰频率变更信息、以及表示各共振峰幅度的变更内容的共振峰幅度变更信息。

在各共振峰频率变更信息和各共振峰幅度变更信息中,分别包含有表示共振峰频率的变换量、振荡速度、振荡幅度的信息以及表示共振峰幅度的变换量、振荡速度、振荡幅度的信息。关于共振峰频率和共振峰幅度的变换量、振荡速度、振荡幅度,将在后面详细说明。

图 8 是表示由音质变更部 250 执行的音质变更处理的流程图。

音质变更部(生成单元)250 在从文本解析部 220 接收到音质数

据编号之后,将该音质数据编号作为检索关键字,对音质数据库 260 进行检索,取得对应的音质参数(步骤 S1)。然后,音质变更部 250 参照包含在所取得的音质参数中的性别变更标记,判断是否应变更合成语音的性别(即是否应执行性别变更处理)(步骤 S2)。在例如性别变更标记被设定为“0”,从而音质变更部 250 判断不应进行性别变更时,跳过步骤 S3,进入步骤 S4,而在例如性别变更标记被设定为“1”,音质变更部 250 判断应进行性别变更时,进入步骤 S3,执行性别变更处理。

图 9 是例示出存储在存储单元(图示省略)中的性别变更处理用的映射函数 mf 的图,图 10 和图 11 是表示男性和女性分别朗读同一音素(例如「あ」等)时的分析结果的图。另外,图 9 所示的映射函数 mf 的横轴表示输入频率(被输入音质变更部 250 的共振峰频率),纵轴表示输出频率(从音质变更部 250 输出的频率变更后的共振峰频率), f_{max} 表示可输入的共振峰频率的最大值。此外,图 10 和图 11 所示的分析图表 $g1$ 、 $g2$ 的横轴表示频率,纵轴表示幅度。

通过比较图 10 和图 11 所示的分析图表 $g1$ 、 $g2$ 可知,男性音素的第 1 共振峰频率 f_{m1} ~第 4 共振峰频率 f_{m4} 比女性音素的第 1 共振峰频率 f_{f1} ~第 4 共振峰频率 f_{f4} 低。因此,在本实施方式中,如图 9 所示,利用位于直线 $n1$ (输入频率=输出频率,参照虚线部分)上侧的映射函数 mf (参照实线部分),来将具有男性特征的音素变更为具有女性特征的音素。

具体地讲,音质变更部 250 利用图 9 所示的映射函数 mf ,将所输入的音素数据的各共振峰频率向频率高的方向变换。由此,所输入的男性音素的各共振峰频率被变更为具有女性特征的共振峰频率。另外,在输入女性音素的共振峰频率的情况下,与上述情况相反,可以利用位于直线 $n1$ 下侧的映射函数 mf' (参照图 9 中点划线所示的部分)。

音质变更部 250 在执行了上述性别变更处理,进入步骤 S4 之后,按照各共振峰变更信息所表示的各共振峰频率的变换量,对各共振峰频率进行变换。进而,音质变更部 250 使变换后的各共振峰频率振荡,

执行频率振荡处理（步骤 S5）。

图 12 是例示出存储在存储单元（图示省略）中的、在频率振荡处理中使用的振荡表 TA 的图，图 13 是例示出从该振荡表 TA 中读出的振荡值和时间之间的关系的图。在本实施方式中，为了说明方便，假定使用同一振荡表 TA 来使上述各共振峰频率振荡的情况，但也可以对于每个共振峰频率使用振荡值等不同的振荡表。

振荡表 TA 是按照时间顺序登录振荡值的表。音质变更部 250 按照各共振峰变更信息所表示的共振峰频率的振荡速度，来控制在振荡表 TA 中登录的振荡值的读出速度（或者跳过（即不读出）振荡值的数量），另一方面，执行频率振荡处理，即将所读出的各振荡值乘以各共振峰变更信息所表示的共振峰频率的振荡幅度。由此，可以获得使图 14 所示的共振峰频率 f_m 以振荡速度 sp 、振荡幅度 lv 振荡的波形。在本实施方式中，为了减少共振峰频率的振荡幅度的运算量，例示了利用上述振荡表 TA 的方式，但也可以不利用振荡表 TA，而是利用规定的函数来求出共振峰频率的振荡幅度。

音质变更部 250 在执行了频率振荡处理之后，进入步骤 S6，按照各共振峰变更信息所表示的各共振峰幅度的变换量，来对各共振峰幅度进行变换。进而，音质变更部 250 使变换后的各共振峰幅度振荡，执行幅度振荡处理（步骤 S7），然后结束处理。另外，对于在幅度振荡处理中使用的振荡表以及利用该振荡表来使各共振峰幅度振荡情况下的动作，由于可以与使上述各共振峰频率振荡的情况大致同样地说明，这里省略其说明。此外，对于共振峰幅度的振荡，可以使用与共振峰频率的振荡同样的振荡表来使其振荡，但也可以使用与共振峰频率的振荡不同的振荡表来使其振荡。

音质变更部（生成单元）250 在根据所取得的音质参数（音素数据加工信息）而对各音素的音质进行变更（即对音素数据进行加工）之后，将每个共振峰的基本波形选择信息、各共振峰频率和各共振峰幅度提供给语音信号生成部 270。

语音信号生成部 270 在接收到从音质变更部 250 提供的基本波形选择信息之后，从波形数据库 280 中取得该基本波形选择信息所表

示的波形数据。该基本波形选择信息所表示的基本波形对于每个共振峰可以不同，例如可以使频率低的共振峰的基本波形为正弦波，而使表现个性的频率高的共振峰的基本波形为正弦波以外的波形（例如矩形波或锯齿波等）等。当然，也可以不利用多种基本波形，而是仅利用单一的基本波形（例如正弦波）。

语音信号生成部（生成单元）270 在按照这样的方式选择了各波形数据之后，利用所选择的各波形数据、各共振峰频率、各共振峰幅度，生成每个共振峰的共振峰波形。然后，语音信号生成部（生成单元）270 对各共振峰波形进行加法运算，生成合成语音信号。然后，语音信号生成部 270 对所生成的合成语音信号进行赋予音高的处理（以下称为音高赋予处理），该音高是从文本解析部 220 提供的音高信息（音高指定信息）所表示的音高。

图 15 是用于说明音高赋予处理的图。在图 15 中，为了容易理解说明，例示出了对正弦波的合成语音信号赋予音高的情况。

语音信号生成部 270 根据从文本解析部 220 提供的音高信息，计算出图 15 所示的时间包络线 t_p 的周期。其中，合成语音的音高依赖于时间包络线 t_p 的周期，时间包络线 t_p 的周期越长，音高越低，而时间包络线的周期越短，音高越高。语音信号生成部 270 在按照这样的方式求出时间包络线 t_p 的周期之后，以所求得的时间包络线 t_p 的周期，反复对时间包络线 t_p 和合成语音信号进行乘法运算，由此得到被赋予了规定音高的合成语音信号。

图 16 是例示出进行了音质变更处理和音高赋予处理之后的特定共振峰的共振峰波形的图。如图 16 所示，与音质变更相关的处理（例如共振峰频率和共振峰幅度的振荡处理等）可以用帧周期（帧单位）进行控制。语音信号生成部（生成单元）270 在得到上述的被赋予了规定音高的合成语音信号之后，将其作为合成语音向外部输出。由此，用户可以通过所希望的音质的合成语音来确认被输入语音合成装置 100 的文本（「こんにちは」等）的内容。

如上所述，根据本实施方式的语音合成装置，由于可以在音质变更部中进行以共振峰为单位的各种音质变更处理，所以即使所存储

的音素数据只有 1 种（即只有特定朗读者的音素数据），也能够进行各种音质的语音合成。

B. 其他

在上述本实施方式中，例示出了在被输入语音合成装置 100 的文本信息中包含音高信息的情况（参照图 2），但也可以在该文本信息中不包含音高信息。假定该情况，当预先在音素数据库 240 中登录替代音高信息（参照图 3 的括号中部分），而在文本信息中不包含音高信息时，可以利用该替代音高信息所表示的音高（例如 C（哆）等）来作为合成语音的音高。此外，除了替代音高信息之外，还可以在音素数据库 240 中预先登录图 4 所示的每个帧的共振峰信息的数量（共振峰数量信息，参照图 3 的括号中部分）。

此外，为了通过 CPU（或 DSP）执行存储在 ROM 等存储器中的程序，以实现以上说明的语音合成装置 100 的各种功能，可以将上述程序记录在 CD-ROM 等记录介质上而颁布，或者可以经由互联网等通信网络来颁布。

在以上的说明中，语音变更处理是基于从文本信息中获得的音质数据编号而进行的，但也可以从所输入的文本信息中自动抽取出关键字，然后利用所抽取出的关键字，来参照预先设置在电子设备内的、具有每个音质的关键字的数据库，由此来自动判定适合于该文本信息的音质。

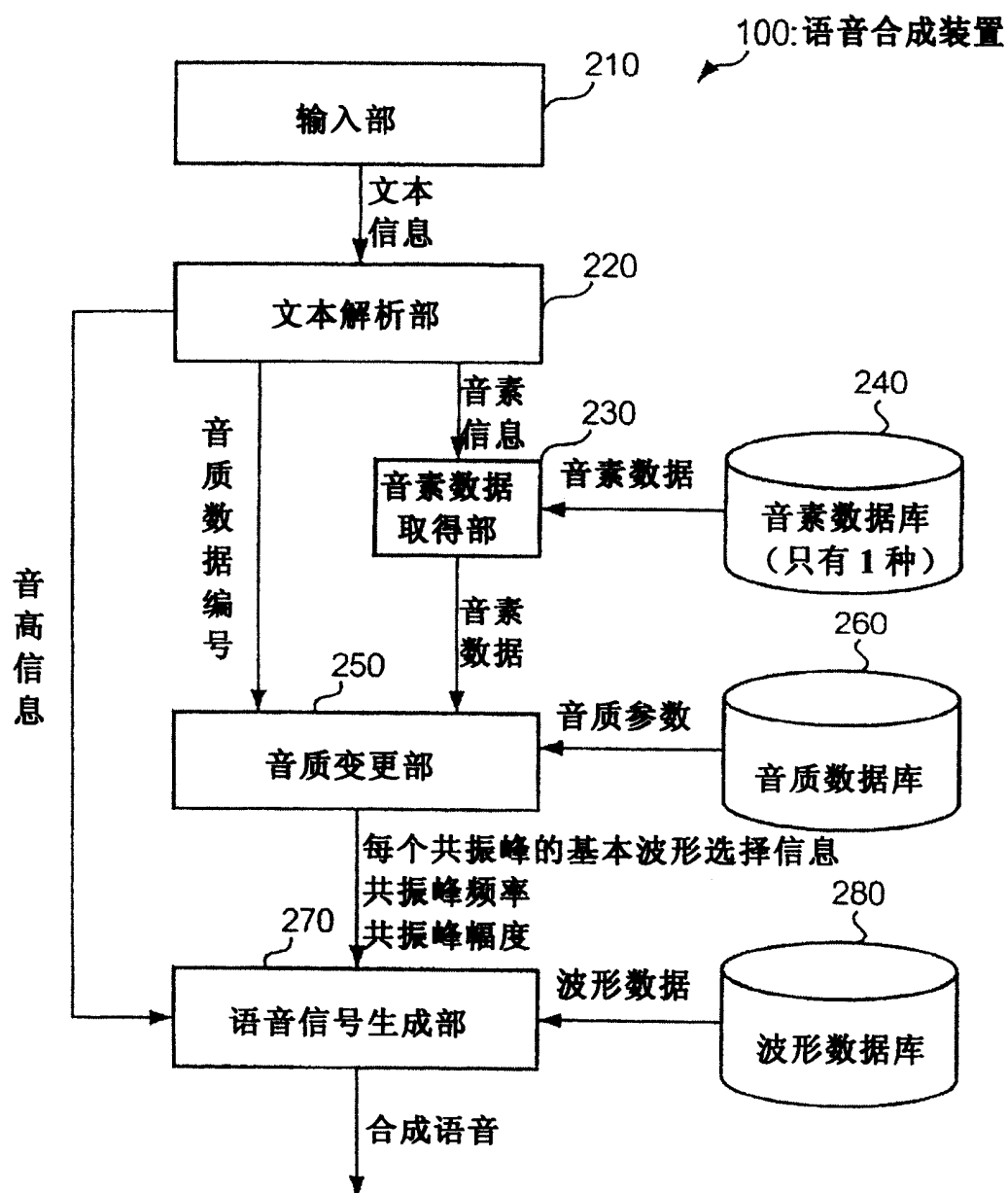


图 1

〈文本信息〉

文本内容信息（「こんにちは」等）
⋮
音质数据编号（K1~Kn 等）
音高信息（C（哆）~B（唏）等）

图 2

240:音素数据库
↙

登录音素数据数量=m
音素数据 1（「あ」等）
音素数据 2（「い」等）
⋮
音素数据 m（「ん」等）
⋮
（替代音高信息）
（共振峰数量信息=k）

图 3

〈音素数据〉

第 1 帧信息	浊音/清音判别标记	
	第 1 共振峰信息	第 1 共振峰频率
		第 1 共振峰幅度
	第 2 共振峰信息	第 2 共振峰频率
		第 2 共振峰幅度
	⋮	⋮
	第 k 共振峰信息	第 k 共振峰频率
		第 k 共振峰幅度
第 2 帧信息	浊音/清音判别标记	
	第 1 共振峰信息	第 1 共振峰频率
		第 1 共振峰幅度
	第 2 共振峰信息	第 2 共振峰频率
		第 2 共振峰幅度
	⋮	⋮
	第 k 共振峰信息	第 k 共振峰频率
		第 k 共振峰幅度
⋮		
第 n 帧信息	⋮	⋮

图 4

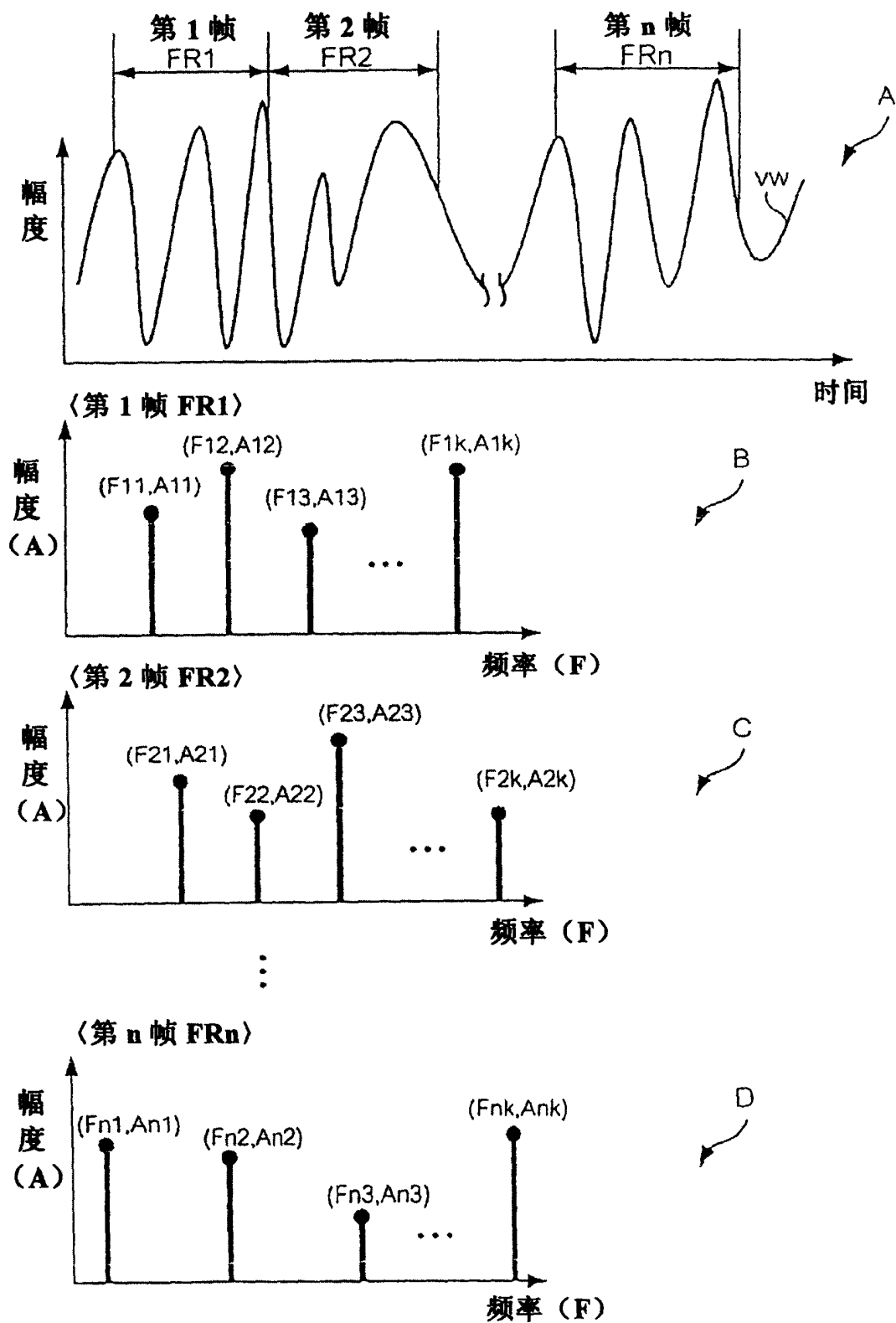


图 5

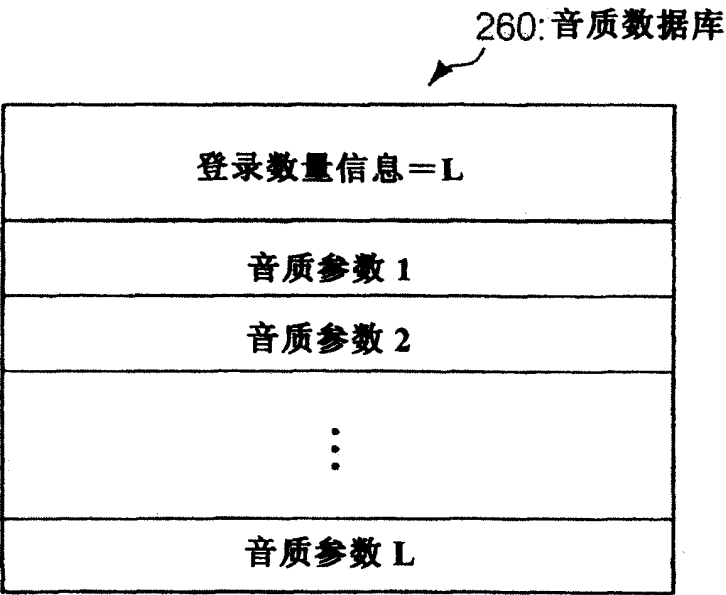


图 6

〈音质参数〉

音质数据编号		
性别变更标记		
第 1 共振峰 变更信息	基本波形选择信息	
	共振峰频率 变更信息	共振峰频率的变换量
		共振峰频率的振荡速度
		共振峰频率的振荡幅度
	共振峰幅度 变更信息	共振峰幅度的变换量
		共振峰幅度的振荡速度
		共振峰幅度的振荡幅度
第 2 共振峰 变更信息	...	...
⋮	⋮	
第 k 共振峰 变更信息	...	...

图 7

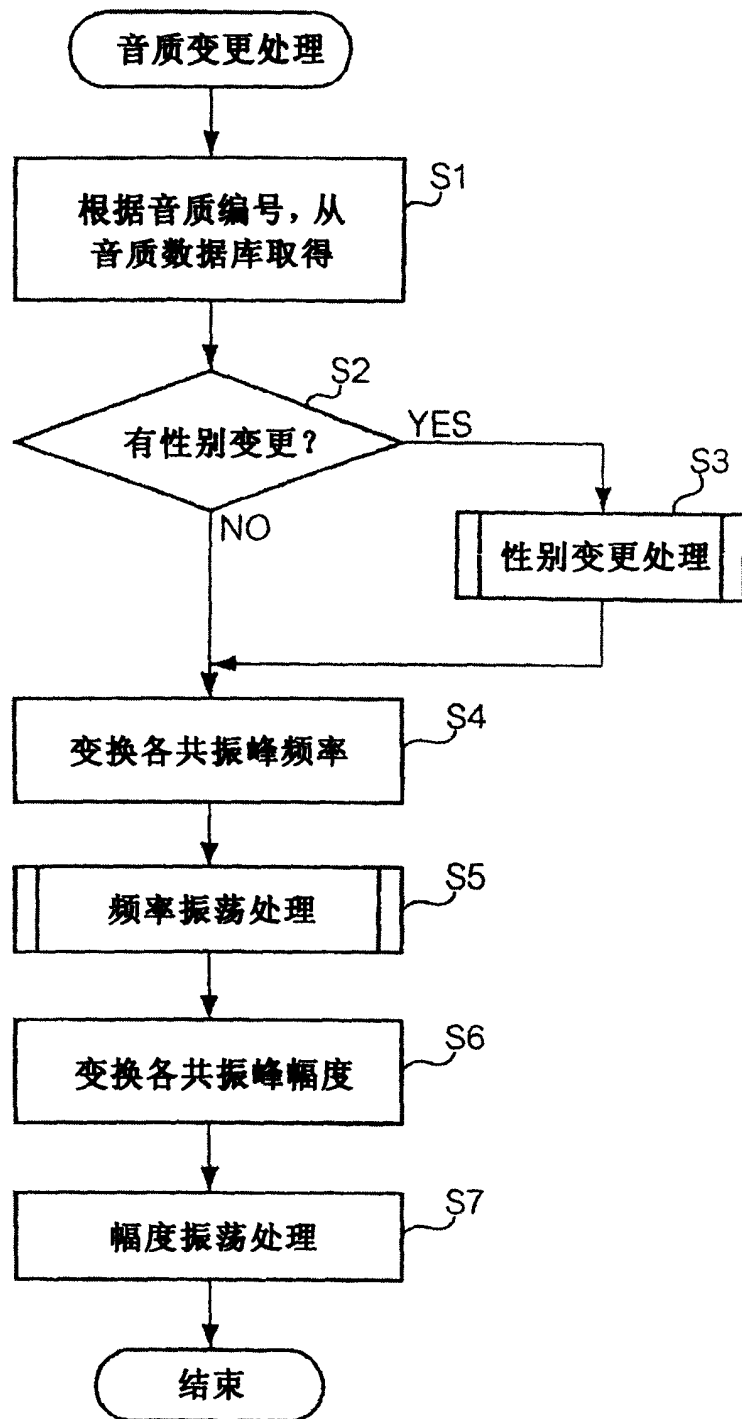


图 8

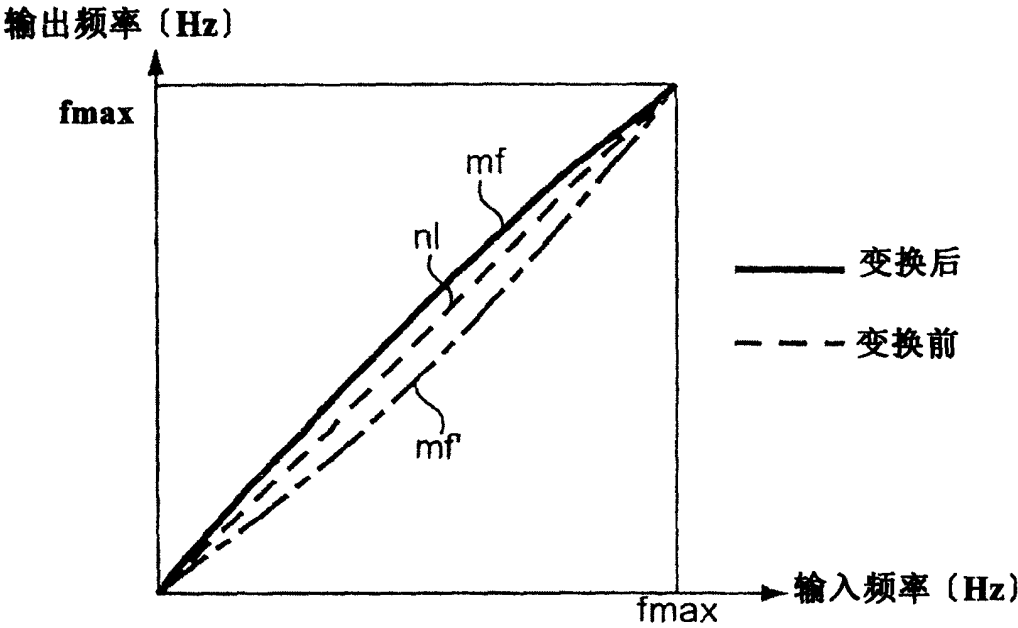


图 9

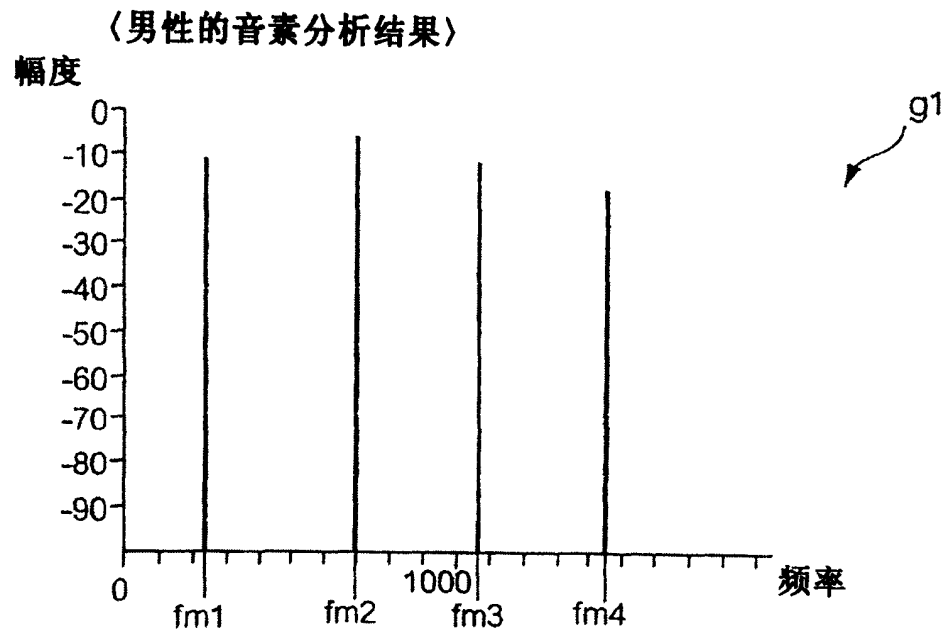


图 10

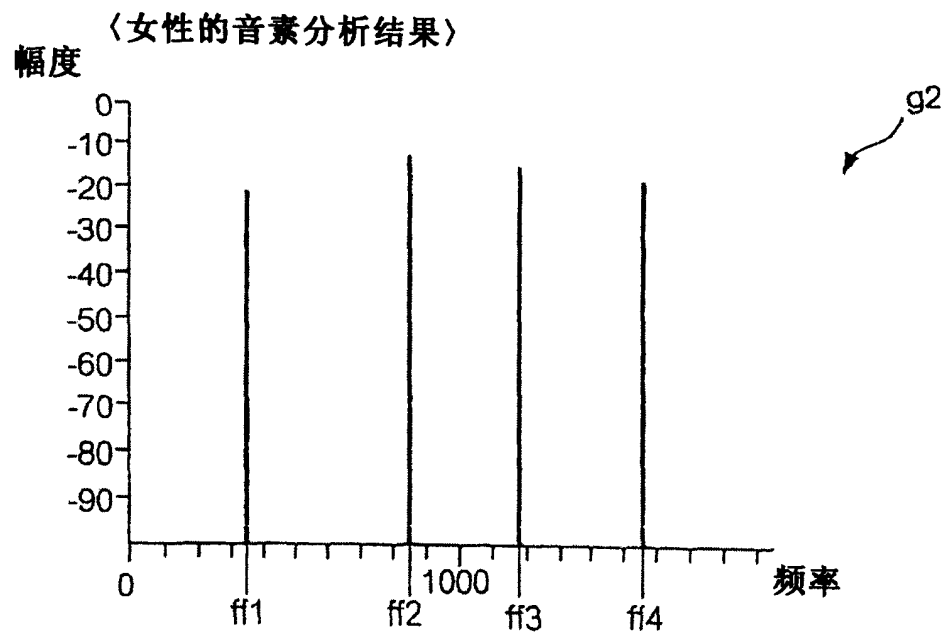


图 11

〈振荡表 TA〉

0	25	50	75	100	75	50	25	0	-25	-50	-75	-100	-75	-50	-25	0
---	----	----	----	-----	----	----	----	---	-----	-----	-----	------	-----	-----	-----	---

图 12

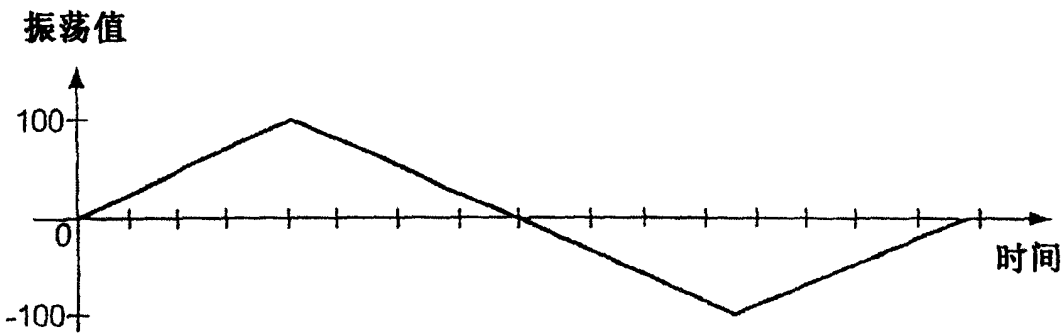


图 13

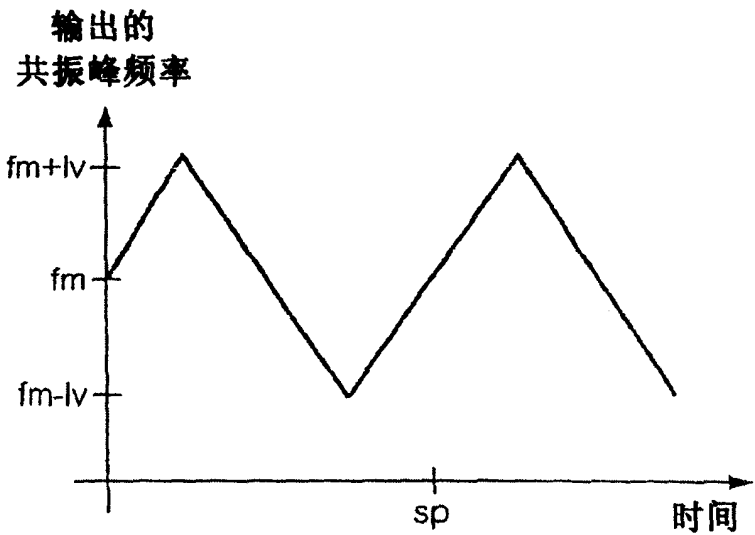


图 14

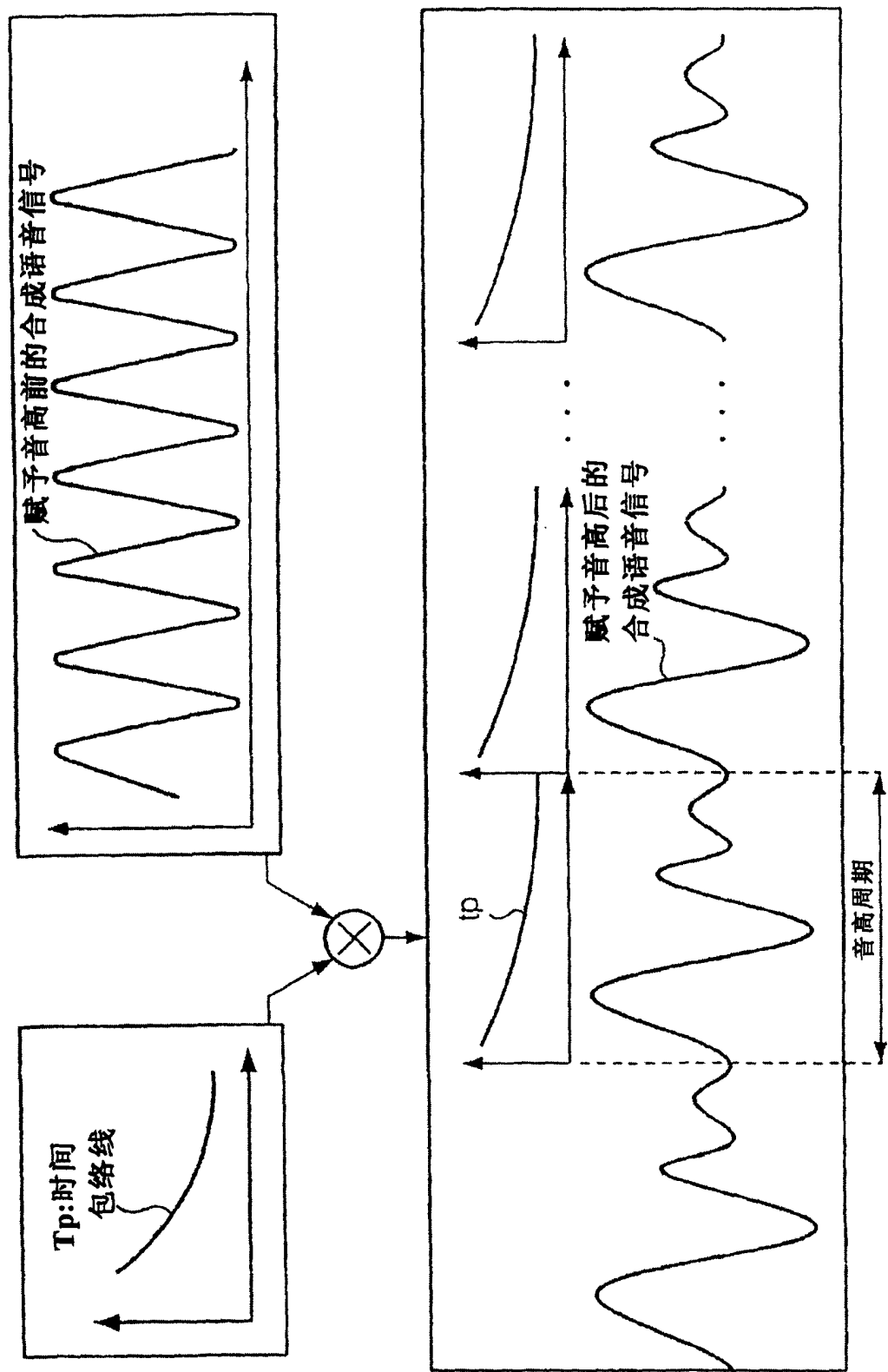


图 15

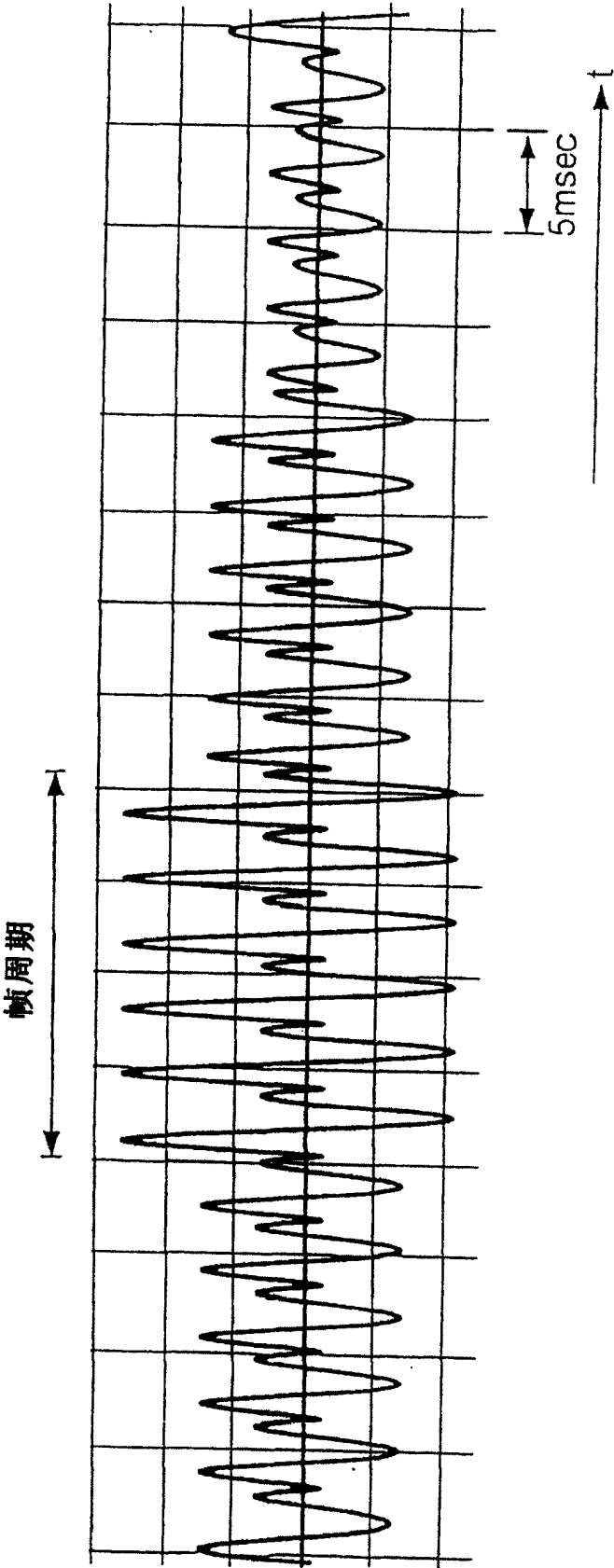


图 16

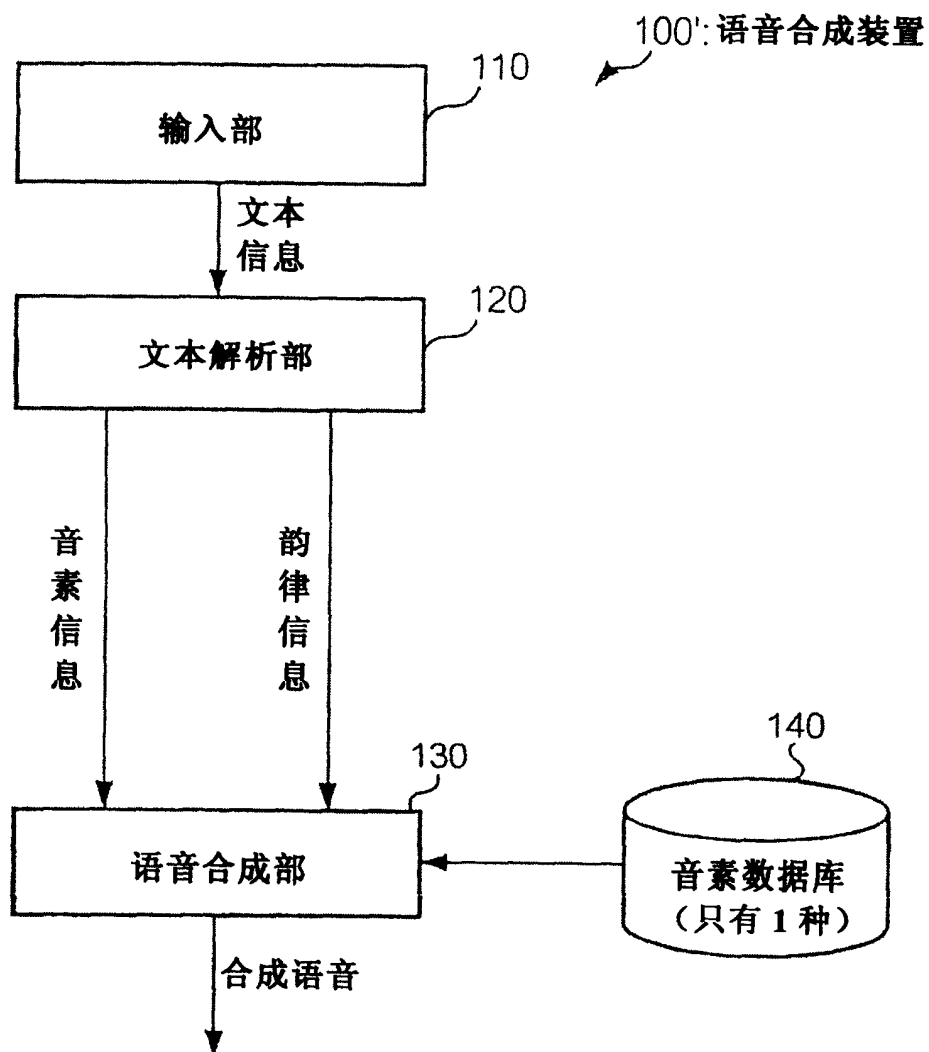


图 17