(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 102822888 A
(43)申请公布日 2012.12.12

(21)申请号 201180016109.9

(22)申请日 2011.03.23

(30)优先权数据
2010-070378 2010.03.25 JP

(85)PCT申请进入国家阶段日
2012.09.25

(86)PCT申请的申请数据
PCT/JP2011/001696 2011.03.23

(87)PCT申请的公布数据
WO2011/118207 JA 2011.09.29

(71)申请人 日本电气株式会社
地址 日本东京都

(72)发明人 加藤正德

(74)专利代理机构 北京市金杜律师事务所 11256
代理人 王茂华 辛鸣

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

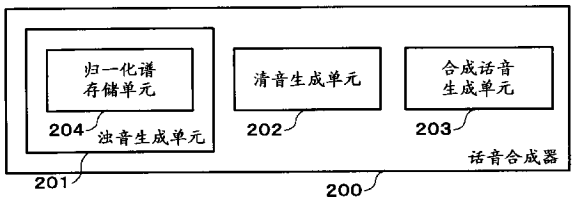
权利要求书 2 页 说明书 12 页 附图 5 页

(54)发明名称

话音合成器、话音合成方法和话音合成程序

(57)摘要

归一化谱存储单元(204)预存储基于随机数序列计算的归一化谱。浊音生成单元(201)基于与输入文本相对应的浊音的多个分段以及存储在归一化谱存储单元(204)中的归一化谱来生成浊音波形。清音生成单元(202)基于与输入文本相对应的清音的多个分段来生成清音波形。合成话音生成单元(203)基于由浊音生成单元(201)生成的浊音波形和由清音生成单元(202)生成的清音波形来生成合成话音。



CN 102822888 A

1. 一种话音合成器，其生成输入文本的合成话音，包括：

浊音生成单元，其包括预存储基于随机数序列而计算的一个或多个归一化谱的归一化谱存储单元，并且基于与所述文本相对应的浊音的多个分段和存储在所述归一化谱存储单元中的所述归一化谱，生成浊音波形；

清音生成单元，其基于与所述文本相对应的清音的多个分段，生成清音波形；以及

合成话音生成单元，其基于由所述浊音生成单元生成的所述浊音波形和由所述清音生成单元生成的所述清音波形，生成所述合成话音。

2. 根据权利要求 1 所述的话音合成器，其中所述浊音生成单元基于存储在所述归一化谱存储单元中的所述归一化谱以及幅度谱，生成多个音高波形作为与所述文本相对应的浊音的分段，并且基于所述生成的音高波形，生成所述浊音波形。

3. 根据权利要求 1 所述的话音合成器，其中所述浊音生成单元基于存储在所述归一化谱存储单元中的所述归一化谱，生成时域波形，基于所述生成的时域波形和与所述输入文本相对应的韵律，生成激励信号，并且基于所述生成的激励信号，生成所述浊音波形。

4. 根据权利要求 1 至 3 中任一项所述的话音合成器，其中通过使用基于随机数序列的群延迟来计算的一个或多个归一化谱被预存储在所述归一化谱存储单元中。

5. 根据权利要求 1 至 4 中任一项所述的话音合成器，其中

所述归一化谱存储单元预存储两个或者更多个归一化谱，并且

所述浊音生成单元通过使用与用于生成先前的浊音波形的归一化谱不同的归一化谱，生成每个浊音波形。

6. 根据权利要求 1 至 5 中任一项所述的话音合成器，其中存储在所述归一化谱存储单元中的归一化谱的数目在 2 至一百万的范围内。

7. 一种用于生成输入文本的合成话音的话音合成方法，包括：

基于与所述文本相对应的浊音的多个分段和存储在用于预存储基于随机数序列而计算的归一化谱的归一化谱存储单元中的一个或多个归一化谱，生成浊音波形；

基于与所述文本相对应的清音的多个分段，生成清音波形；以及

基于所述生成的浊音波形和所生成的清音波形，生成所述合成话音。

8. 根据权利要求 7 所述的话音合成方法，其中

基于存储在所述归一化谱存储单元中的所述归一化谱和幅度谱，生成多个音高波形作为与所述文本相对应的浊音的分段，并且

基于所述生成的音高波形来生成所述浊音波形。

9. 一种待安装在话音合成器中的话音合成程序，所述话音合成器生成输入文本的合成话音，其中所述话音合成程序使得计算机执行：

浊音波形生成过程，所述浊音波形生成过程基于与所述文本相对应的浊音的多个分段以及存储在用于预存储基于随机数序列而计算的归一化谱的归一化谱存储单元中的一个或多个归一化谱，生成浊音波形；

清音波形生成过程，所述清音波形生成过程基于与所述文本相对应的清音的多个分段，生成清音波形；并且

合成话音生成过程，所述合成话音生成过程基于在所述浊音波形生成过程中生成的所述浊音波形以及在所述清音波形生成过程中生成的所述清音波形，生成所述合成话音。

10. 根据权利要求 9 所述的话音合成程序，其中所述浊音波形生成过程基于存储在所述归一化谱存储单元中的所述归一化谱以及幅度谱，生成多个音高波形作为与所述文本相对应的浊音的分段，并且基于所述生成的音高波形，生成所述浊音波形。

话音合成器、话音合成方法和话音合成程序

技术领域

[0001] 本发明涉及用于生成输入文本的合成话音的话音合成器、话音合成方法和话音合成程序。

背景技术

[0002] 存在通过基于由文本分析的结果表示的语音信息的规则、借助于话音合成而分析文本以及生成合成话音的话音合成器。

[0003] 这种通过规则、借助于话音合成而生成合成话音的话音合成器首先基于文本的分析的结果而生成关于合成话音的韵律信息（通过声音的音高（音高频率）、声音的长度（音位持续时间）、声音的量级（功率）等来指示韵律的信息）。随后，话音合成器从分段词典中选择与文本分析的结果和韵律信息相对应的分段（合成单元），该分段词典已经预存储了多种分段（波形生成参数）。

[0004] 随后，话音合成器基于从分段词典中选择的分段（波形生成参数）来生成话音波形。最后，话音合成器通过连接所生成的话音波形来生成合成话音。

[0005] 当此类话音合成器基于所选择的分段来生成话音波形时，话音合成器生成具有与由所生成的韵律信息所指示的韵律接近的韵律的话音波形，以便生成高声音质量的合成话音。

[0006] 非专利文献 1 描述了一种用于生成话音波形的方法。在非专利文献 1 的方法中，将振幅谱（作为通过对音频信号进行傅里叶变换而获得的谱的振幅分量）在时间频率方向进行平滑，并且将其用作波形生成参数。非专利文献 1 还描述了一种用于将归一化谱计算为通过振幅谱进行归一化的谱的方法。在该方法中，基于随机数来计算群延迟，并且通过使用所计算的群延迟来计算归一化谱。

[0007] 专利文献 1 描述了一种话音处理设备，包括存储单元，该存储单元预存储待用于生成合成话音的过程的话音分段波形的周期分量和非周期分量。

[0008] 引用列表

[0009] 专利文献

[0010] 专利文档 1 :JP-A-2009-163121(0025-0289 段，图 1)

[0011] 非专利文献

[0012] 非专利文献 1 :Hideki Kawahara, ″ Speech Representation and Transformation Using Adaptive Interpolation of Weighted Spectrum :Vocoder Revisited″ ,(USA),IEEE ICASSP-97,第 2 卷，1997，第 1303-1306 页

发明内容

[0013] 技术问题

[0014] 在由前述话音合成器采用的波形生成方法中，连续地计算归一化谱。归一化谱用于生成音高波形，其必须以接近音高周期的间隔生成。因此，采用该波形生成方法的话音合

成器必须频繁地计算归一化谱，从而引起极大量的计算。

[0015] 另外，归一化谱的计算需要如非专利文献 1 中所描述的基于随机数的群延迟的计算。在通过使用群延迟来计算归一化谱的过程中，必须执行包括大量计算的积分计算。由此，采用上述波形生成方法的话音合成器必须频繁地执行一系列计算（基于随机数的群延迟的计算以及通过进行包括大量计算的积分计算而通过所计算的群延迟的对归一化谱的计算）。

[0016] 随着计算数量的增多，话音合成器生成合成话音所需要的吞吐量（每单位时间的工作负载）增加。因此，每单位时间应当输出的合成话音的生成变得不可能，尤其是在低处理功率的话音合成器与合成话音的生成同步地输出合成话音时。平滑输出合成话音的不可能性严重地影响了由话音合成器输出的合成话音的声音质量。

[0017] 同时，专利文献 1 中描述的话音处理设备通过使用存储单元中预存储的话音分段波形的周期分量和非周期分量来生成合成话音。需要此类话音处理设备来生成更高声音质量的合成话音。

[0018] 因此，本发明的主要目的是提供一种话音合成器、话音合成方法和话音合成程序，其能够利用较少数目的计算来生成更高声音质量的合成话音。

[0019] 问题的解决方案

[0020] 为了实现上述目的，本发明提供了一种话音合成器，该话音合成器生成输入文本的合成话音，包括 ：浊音生成单元，其包括预存储基于随机数序列而计算的一个或多个归一化谱的归一化谱存储单元，并且基于与文本相对应的浊音的多个分段和存储在归一化谱存储单元中的归一化谱来生成浊音波形 ；清音生成单元，其基于与文本相对应的清音的多个分段来生成清音波形 ；以及合成话音生成单元，其基于由浊音生成单元生成的浊音波形和由清音生成单元生成的清音波形来生成合成话音。

[0021] 本发明还提供了一种话音合成方法，用于生成输入文本的合成话音，包括 ：基于与文本相对应的浊音的多个分段和存储在用于预存储基于随机数序列而计算的归一化谱的归一化谱存储单元中的一个或多个归一化谱来生成浊音波形 ；基于与文本相对应的清音的多个分段来生成清音波形 ；以及，基于所生成的浊音波形和所生成的清音波形来生成合成话音。

[0022] 本发明还提供了一种待安装在话音合成器中的话音合成程序，该话音合成器生成输入文本的合成话音，其中该话音合成程序使得计算机执行 ：浊音波形生成过程，该浊音波形生成过程基于与文本相对应的浊音的多个分段以及存储在用于预存储基于随机数序列而计算的归一化谱的归一化谱存储单元中的一个或多个归一化谱来生成浊音波形 ；清音波形生成过程，该清音波形生成过程基于与文本相对应的清音的多个分段来生成清音波形 ；以及，合成话音生成过程，该合成话音生成过程基于在浊音波形生成过程中生成的浊音波形以及在清音波形生成过程中生成的清音波形来生成合成话音。

[0023] 本发明的有益效果

[0024] 根据本发明，通过使用预存储在归一化谱存储单元中的归一化谱来生成合成话音的波形。因此，在生成合成话音时可以省略归一化谱的计算。从而，可以减少在话音合成时必需的计算的数目。

[0025] 另外，由于归一化谱用于生成合成话音波形，所以与话音分段波形的周期分量和

非周期分量用于生成合成话音的情况相比，可以生成更高声音质量的合成话音。

附图说明

[0026] [图 1] 其绘出了示出根据本发明的第一示例性实施方式的话音合成器的配置的示例的框图。

[0027] [图 2] 其绘出了示出由目标分段环境指示的每条信息和由关于候选分段 A1 和 A2 的属性信息指示的每条信息的表。

[0028] [图 3] 其绘出了示出由关于候选分段 A1、A2、B1 和 B2 的属性信息指示的每条信息的表。

[0029] [图 4] 其绘出了示出用于计算待存储在归一化谱存储单元中的归一化谱的过程的流程图。

[0030] [图 5] 其绘出了示出第一示例性实施方式中的话音合成器的波形生成单元的操作的流程图。

[0031] [图 6] 其绘出了示出根据本发明的第二示例性实施方式的话音合成器的配置的示例的框图。

[0032] [图 7] 其绘出了示出第二示例性实施方式中的话音合成器的波形生成单元的操作的流程图。

[0033] [图 8] 其绘出了示出根据本发明的话音合成器的主体部分的框图。

具体实施方式

[0034] <第一示例性实施方式>

[0035] 以下将参考附图描述根据本发明的话音合成器的第一示例性实施方式。图 1 是示出根据本发明的第一示例性实施方式的话音合成器的配置的示例的框图。

[0036] 如图 1 中所示，根据本发明的第一示例性实施方式的话音合成器包括波形生成单元 4。波形生成单元 4 包括浊音生成单元 5、清音生成单元 6 和波形连接单元 7。如图 1 中所示，波形生成单元 4 经由分段选择单元 3 和韵律生成单元 2 连接至语言处理单元 1。分段信息存储单元 12 连接至分段选择单元 3。

[0037] 如图 1 中所示，浊音生成单元 5 包括归一化谱存储单元 101、归一化谱加载单元 102、傅里叶逆变换单元 55 和音高波形叠加单元 56。

[0038] 分段信息存储单元 12 已经存储了分别针对各话音合成单元而生成的分段（话音分段）以及关于每个分段的属性信息。分段例如是针对每个话音合成单元而分段（剪切、提取）的话音波形、从分段的话音波形中提取的波形生成参数（线性预测分析参数、倒谱系数等）的时间序列等。将采用浊音的分段是幅度谱而清音的分段是分段（剪切、提取）的话音波形的情况的示例来给出以下说明。

[0039] 关于分段的属性信息包括音韵信息（指示声音（话音）的音素环境、音高频率、幅度、持续时间等作为每个分段的基础）和韵律信息。在很多情况下，从由人发出的语音（自然话音波形）中提取或者生成分段。例如，有时从由广播员或者配音演员发出的语音的所记录的声音数据中提取或者生成分段。

[0040] 发出作为分段的基础的语音的人（说话者）称为分段的“原始说话者”。音素、音

节、半音节 (demisyllable)(例如,CV(C :辅音,V :元音))、CVC、VCV 等通常被用作话音合成单元。

[0041] 以下参考文献 1 和参考文献 2 包括对合成单元和分段的长度的说明。

[0042] 参考文献 1 :Huang, Acero, Hon, " Spoken Language Processing, " Prentice Hall,2001,第 689-836 页

[0043] 参考文献 2 :Masanobu Abe 等人, " An Introduction to Speech Synthesis Units, " IEICE(电子、信息和通信工程师协会(日本))技术报告,第 100 卷,No.392,2000,第 35-42 页

[0044] 语言处理单元 1 分析输入文本的文字。具体地,语言处理单元 1 执行诸如形态分析、解析或者阅读分析之类的分析。基于分析的结果,语言处理单元 1 向韵律生成单元 2 和分段选择单元 3 输出指示表示“阅读”的符号串(例如,音素符号)的信息和指示每个词素的话音、词形变化、口音类型等的部分的信息,作为语言分析结果。

[0045] 韵律生成单元 2 基于由语言处理单元 1 输出的语言分析结果来生成合成话音的韵律。韵律生成单元 2 向分段选择单元 3 和波形生成单元 4 输出指示所生成的韵律的韵律信息,作为目标韵律信息(目标韵律学信息)。通过在以下参考文献 3 中描述的方法生成韵律,例如 :

[0046] 参考文献 3 :Yasushi Ishikawa, " Prosodic Control for Japanese Text-to-Speech Synthesis, " IEICE(电子、信息和通信工程师协会(日本))技术报告,第 100 卷,No.392,2000,第 27-34 页

[0047] 分段选择单元 3 基于语言分析结果和目标韵律信息从存储在分段信息存储单元 12 中的分段中选择满足规定条件的分段。分段选择单元 3 向波形生成单元 4 输出所选择的分段和关于分段的属性信息。

[0048] 以下将说明用于从存储在分段信息存储单元 12 中的分段中选择满足规定条件的分段的分段选择单元 3 的操作。基于输入的语言分析结果和目标韵律信息,分段选择单元 3 生成针对每个话音合成单元的指示合成话音的特性的信息(在下文中称为“目标分段环境”)。

[0049] 目标分段环境是包括以下内容的信息 :有关音素(构成作为目标分段环境的生成的目标的合成话音)、在前音素(作为有关音素之前的音素)、在后音素(作为有关音素之后的音素)、重音存在 / 不存在、与口音调核 (accent nucleus) 的距离、每个话音合成单元的音高频率、功率、每个话音合成单元的持续时间、倒谱、MFCC(美尔频率倒谱系数)、这些值的 Δ 量(每单位时间的变化)等。

[0050] 随后,针对每个话音合成单元,分段选择单元 3 基于包括在所生成的目标分段环境中的信息而从分段信息存储单元 12 获取与连续音素相对应的多个分段。具体地,分段选择单元 3 基于包括在目标分段环境中的信息而从分段信息存储单元 12 中获取与有关音素相对应的多个分段、与在前音素相对应的多个分段以及与在后音素相对应的多个分段。所获取的分段是用于生成合成话音的分段的候选(在下文中,称为“候选分段”)。

[0051] 继而,针对相邻候选分段(例如,与有关音素相对应的候选分段和与在前音素相对应的候选分段)的每个组合,分段选择单元 3 计算“成本”作为表示组合作为用于生成语音(话音)的分段的适用性程度的指数。成本是目标分段环境和关于每个候选分段的属性

信息之间的差异以及相邻候选分段之间的属性信息的差异的计算的结果。

[0052] 成本（计算结果的值）随着合成话音的特性（由目标分段环境表示）与候选分段之间的相似度的增大而降低，也即，随着用于生成语音（话音）的组合的适用性程度的增大而降低。随着被使用的分段的成本的降低，指示与由人发出的话音的相似性程度的合成话音的自然度增加。分段选择单元 3 选择所计算的成本最低的分段。

[0053] 具体地，由分段选择单元 3 计算的成本包括单位成本和连接成本。单位成本指示当候选分段在由目标分段环境表示的环境中使用时，推测发生的声音质量恶化的程度。基于关于候选分段的属性信息和目标分段环境之间的相似性程度来计算单位成本。

[0054] 连接成本指示推测由于连接的话音分段之间的分段环境的不连续性而发生的声音质量恶化的程度。基于相邻候选分段之间的分段环境的亲和度来计算连接成本。已经提出了用于计算单位成本和连接成本的各种方法。

[0055] 通常，通过使用包括在目标分段环境中的信息来计算单位成本。通过使用以下项来计算连接成本：相邻分段的连接边界处的音高频率、倒谱、MFCC、短期自相关、功率、这些值的 Δ 量等。具体地，通过使用从关于分段的多种信息（音高频率、倒谱、功率等）中选择的多条信息来计算单位成本和连接成本。

[0056] 以下将说明计算单位成本的一个示例。图 2 是示出由目标分段环境指示的每条信息以及由关于候选分段 A1 和 A2 的属性信息指示的每条信息的表。

[0057] 在图 2 中所示的示例中，由目标分段环境指示的音高频率是 pitch0[Hz]。由目标分段环境指示的持续时间是 dur0[sec]。由目标分段环境指示的功率是 pow0[dB]。由目标分段环境指示的与口音调核的距离是 pos0。由与候选分段 A1 有关的属性信息指示的音高频率是 pitch1[Hz]。由关于候选分段 A1 的属性信息指示的持续时间是 dur1[sec]。由关于候选分段 A1 的属性信息指示的功率是 pow1[dB]。由关于候选分段 A1 的属性信息指示的与口音调核的距离是 pos 1。类似地，由关于候选分段 A2 的属性信息指示的音高频率、持续时间、功率和与口音调核的距离是 pitch2[Hz]、dur2[sec]、pow2[dB] 和 pos2。

[0058] 附带地，“与口音调核的距离”意味着话音合成单元中与作为口音调核的音素的距离。例如，当在包括 5 个音素的话音合成单元中，第三个音素是口音调核时，与第一音素相对应的分段的“与口音调核的距离”是“-2”。与第二音素相对应的分段的“与口音调核的距离”是“-1”。与第三因素相对应的分段的“与口音调核的距离”是“0”。与第四音素相对应的分段的“与口音调核的距离”是“+1”。与第五音素相对应的分段的“与口音调核的距离”是“+2”。

[0059] 用于计算候选分段 A1 的单位成本 (unit score(A1)) 的公式是：

[0060] unit_score(A1) = (w1×(pitch0-pitch1)^2)

[0061] +(w2×(dur0-dur1)^2)

[0062] +(w3×(pow0-pow1)^2)

[0063] +(w4×(pos0-pos1)^2)

[0064] 用于计算候选分段 A2 的单位成本 (unit_score(A2)) 的公式是：

[0065] unit_score(A2) = (w1×(pitch0-pitch2)^2)

[0066] +(w2×(dur0-dur2)^2)

[0067] +(w3×(pow0-pow2)^2)

[0068] +(w4×(pos0-pos2)^2)

[0069] 在以上公式中，w1-w4 表示预置加权因子。符号“^”表示幂。例如，“2^2”表示 2 的二次幂。

[0070] 以下将说明计算连接成本的示例。图 3 是示出了由关于候选分段 A1、A2、B1 和 B2 的属性信息指示的每条信息的表。附带地，候选分段 B1 和 B2 是针对在具有候选分段 A1 和 A2 作为其候选分段的分段之后的分段的候选分段。

[0071] 在图 3 中所示的示例中，候选分段 A1 的开始边音高频率是 pitch_beg1[Hz]，候选分段 A1 的结束边音高频率是 pitch_end1[Hz]，候选分段 A1 的开始边功率是 pow_beg1[dB]，并且候选分段 A1 的结束边功率是 pow_end1[dB]。候选分段 A2 的开始边音高频率是 pitch_beg2[Hz]，候选分段 A2 的结束边音高频率是 pitch_end2[Hz]，候选分段 A2 的开始边功率是 pow_beg2[dB]，并且候选分段 A2 的结束边功率是 pow_end2[dB]。

[0072] 类似地，候选分段 B 1 的开始边音高频率、结束边音高频率、开始边功率和结束边功率是 pitch_beg3[Hz]、pitch_end3[Hz]、pow_beg3[dB] 和 pow_end3[dB]，并且候选分段 B2 的是 pitch_beg4[Hz]、pitch_end4[Hz]、pow_beg4[dB] 和 pow_end4[dB]。

[0073] 用于计算候选分段 A1 和 B1 的连接成本 (concat_score(A1,B1)) 的公式是：

[0074] concat_score(A1,B1) =

[0075] (c1×(pitch_end1-pitch_beg3)^2)

[0076] +(c2×(pow_end1-pow_beg3)^2)

[0077] 用于计算候选分段 A1 和 B2 的连接成本 (concat_score(A1,B2)) 的公式是：

[0078] concat_score(A1,B2) =

[0079] (c1×(pitch_end1-pitch_beg4)^2)

[0080] +(c2×(pow_end1-pow_beg4)^2)

[0081] 用于计算候选分段 A2 和 B1 的连接成本 (concat_score(A2,B1)) 的公式是：

[0082] concat_score(A2,B1) =

[0083] (c1×(pitch_end2-pitch_beg3)^2)

[0084] +(c2×(pow_end2-pow_beg3)^2)

[0085] 用于计算候选分段 A2 和 B2 的连接成本 (concat_score(A2,B2)) 的公式是：

[0086] concat_score(A2,B2) =

[0087] (c1×(pitch_end2-pitch_beg4)^2)

[0088] +(c2×(pow_end2-pow_beg4)^2)

[0089] 在以上公式中，c1 和 c2 表示预置加权因子。

[0090] 基于所计算的单位成本和连接成本，分段选择单元 3 计算候选分段 A1 和 B1 的组合的成本。具体地，将候选分段 A1 和 B1 的组合的成本计算为 unit(A1)+unit(B1)+concat_score(A1,B1)。同时，将候选分段 A2 和 B1 的组合的成本计算为 unit(A2)+unit(B1)+concat_score(A2,B1)。

[0091] 类似地，将候选分段 A1 和 B2 的组合的成本计算为 unit(A1)+unit(B2)+concat_score(A1,B2)，并且将候选分段 A2 和 B2 的组合的成本计算为 unit(A2)+unit(B2)+concat_score(A2,B2)。

[0092] 分段选择单元 3 从候选分段中选择最小化所计算的成本的分段的组合，作为最适

于语音（话音）的合成的分段。由分段选择单元 3 选择的分段在下文中将被称为“选择的分段”。

[0093] 波形生成单元 4 基于由韵律生成单元 2 输出的目标韵律信息、由分段选择单元 3 输出的分段以及关于分段的属性信息，生成具有与目标韵律信息相一致或者类似的韵律的话音波形。波形生成单元 4 通过连接所生成的话音波形来生成合成话音。由波形生成单元 4 根据分段而生成的话音波形在下文中将被称为“分段波形”，以便使其区别于普通的话音波形。

[0094] 可以将由分段选择单元 3 输出的分段分类为由浊音构成的和由清音构成的。针对浊音的韵律控制所采用的方法与针对清音的韵律控制所采用的方法彼此不同。波形生成单元 4 包括浊音生成单元 5、清音生成单元 6 和波形连接单元 7，该波形连接单元 7 用于连接浊音和清音。分段选择单元 3 向浊音生成单元 5 输出浊音的分段（浊音分段），同时向清音生成单元 6 输出清音的分段（清音分段）。将由韵律生成单元 2 输出的韵律信息输入到浊音生成单元 5 和清音生成单元 6 二者中。

[0095] 基于由分段选择单元 3 输出的清音的分段，清音生成单元 6 生成具有与由韵律生成单元 2 输出的韵律信息相一致或者类似的韵律的清音波形。在该示例中，由分段选择单元 3 输出的清音的分段是分段（剪切、提取）的话音波形。因此，清音生成单元 6 能够通过使用在以下参考文献 4 中描述的方法来生成清音波形：备选地，清音生成单元 6 还可以通过使用在以下参考文献 5 中描述的方法来生成清音波形：

[0096] 参考文献 4：Ryuji Suzuki，Masayuki Misaki，″ Time-scale Modification of Speech Signals Using Cross-correlation，″ (USA)，IEEE 消费电子学报，第 38 卷，1992，第 357-363 页

[0097] 参考文献 5：Nobumasa Seiyama 等人，″ Development of a High-quality Real-time Speech Rate Conversion System，″ 电子、信息和通信工程师协会学报 (Japan)，第 J84-D-2 卷，No. 6，2001，第 918-926 页

[0098] 浊音生成单元 5 包括归一化谱存储单元 101、归一化谱加载单元 102、傅里叶逆变换单元 55 和音高波形叠加单元 56。

[0099] 此处，将给出对谱、幅度谱和归一化谱的说明。谱由特定信号的傅里叶变换定义。在以下参考文献 6 中给出了谱和傅里叶变换的详细说明：

[0100] 参考文献 6：Shuzo Saito，Kazuo Nakata，″ Basics of Phonetical Information Processing″，Ohmsha，Ltd.，1981，第 15-31，73-76 页

[0101] 如参考文献 6 中所述，每个谱由复数表示，并且谱的幅度分量称为“幅度谱”。在该示例中，通过使用其幅度谱对谱进行归一化的结果称为“归一化谱”。当谱由 X(w) 表示时，幅度谱和归一化谱可以在数学上分别表示为 |X(w)| 和 X(w)/|X(w)|。

[0102] 归一化谱存储单元 101 存储先前已经计算的归一化谱。图 4 是示出用于计算待存储在归一化谱存储单元 101 中的归一化谱的过程的流程图。

[0103] 如图 4 中所示，首先生成随机数序列（步骤 S1-1）。基于生成的随机数序列，通过在非专利文献 1 中描述的方法来计算谱的相位分量的群延迟（步骤 S1-2）。在以下参考文献 7 中描述了谱的相位分量和相位分量的群延迟的定义：

[0104] 参考文献 7：Hideki Banno 等，″ Speech Manipulation Method Using Phase

Manipulation Based on Time-Domain Smoothed Group Delay, ″ 电子、信息和通信工程师协会学报 (Japan),第 J83-D-2 卷,No.11,2000,第 2276-2282 页

[0105] 随后,通过使用所计算的群延迟来计算归一化谱(步骤 S1-3)。用于通过使用群延迟来计算归一化谱的方法在参考文献 7 中进行了描述。最后,检查所计算的归一化谱的数目是否已经达到预置数目(设置值)(步骤 S1-4)。如果所计算的归一化谱的数目已经达到预置数目,则该过程结束,否则该过程返回至步骤 S1-1。

[0106] 在步骤 S1-4 中用于检查的预置数目(设置值)等于存储在归一化谱存储单元 101 中的归一化谱的数目。可期望的是,基于随机数序列生成待存储在归一化谱存储单元 101 中的归一化谱,并且生成和存储大量的归一化谱以便保证高随机性。然而,归一化谱存储单元 101 需要具有与归一化谱的数目相对应的高存储量。由此,可期望将在步骤 S1-4 中用于检查的设置值(预置数目)设置为与话音合成器中可允许的最大存储量相对应的最大值。具体地,从声音质量的角度看,如果最多接近一百万的归一化谱存储在归一化谱存储单元 101 中,那么是足够的。

[0107] 另外,存储在归一化谱存储单元 101 中的归一化谱的数目应当是两个或者更多。如果数目是 1,也即,如果仅有一个归一化谱已存储在归一化谱存储单元 101 中,则归一化谱加载单元 102 仅加载一种类型的归一化谱,也即,每次加载相同的归一化谱。在这种情况下,所生成的合成话音的谱的相位分量变为总是不变的,并且不变的相位分量造成声音质量的退化。为此,归一化谱存储单元 101 应当存储两个或者更多个归一化谱。

[0108] 如以上所说明的,在归一化谱存储单元 101 中存储的归一化谱的数目应当设置在 2 至一百万的范围内。由于以下原因,期望存储在归一化谱存储单元 101 中的归一化谱尽可能彼此不同:在归一化谱加载单元 102 以随机顺序从归一化谱存储单元 101 加载归一化谱的情况下,由归一化谱加载单元 102 连续加载相同归一化谱的概率随着存储在归一化谱存储单元 101 中的相同归一化谱的数目的增加而增加。

[0109] 期望存储在归一化谱存储单元 101 中的所有归一化谱之中的相同归一化谱的比率(百分比)低于 10%。如果由归一化谱加载单元 102 连续加载相同的归一化谱,则会如上所述发生由于不变的相位分量而造成的声音质量退化。

[0110] 在归一化谱存储单元 101 中,已经按照随机顺序存储了如下归一化谱,这些归一化谱中的每一个基于随机数序列而生成。为了防止归一化谱加载单元 102 在归一化谱的加载中连续加载相同的归一化谱,期望将归一化谱存储单元 101 内的数据排列为避免在连续的位置处存储相同的归一化谱。利用这样的配置,当由归一化谱加载单元 102 进行归一化谱的连续加载(顺序读取)时,可以防止连续加载两个或者更多个相同归一化谱。

[0111] 另外,为了在由归一化谱加载单元 102 进行归一化谱的随机加载(随机读取)时防止连续使用两个或者更多个相同归一化谱,期望将话音合成器按照如下配置。归一化谱加载单元 102 包括存储装置,该存储装置用于存储已经加载的归一化谱。归一化谱加载单元 102 判断当前过程中加载的归一化谱是否与在先前过程中已经加载并且存储在存储装置中的归一化谱相同。当在当前过程中加载的归一化谱与在先前过程中加载并且存储在存储装置中的归一化谱不同时,归一化谱加载单元 102 利用在当前过程中加载的归一化谱来更新存储在存储装置中的归一化谱。相反,当在当前过程中加载的归一化谱与在先前过程中加载并且存储在存储装置中的归一化谱相同时,归一化谱加载单元 102 重复加载归一化

谱的过程，直到加载了与在先前过程中加载并且存储在存储装置中的归一化谱不同的归一化谱。

[0112] 以下将参考附图说明根据第一示例性实施方式的话音合成器的波形生成单元 4 的操作。图 5 是示出第一示例性实施方式中的话音合成器的波形生成单元 4 的操作的流程图。

[0113] 归一化谱加载单元 102 加载存储在归一化谱存储单元 101 中的归一化谱（步骤 S2-1）。随后，归一化谱加载单元 102 向傅里叶逆变换单元 55 输出加载的归一化谱（步骤 S2-2）。

[0114] 在步骤 S2-1 中，如果归一化谱加载单元 102 按照随机顺序加载归一化谱而不是从归一化谱存储单元 101 的前端（第一地址）依次地进行加载（例如，以存储区域中的地址的顺序），则随机性增加。由此，通过使得归一化谱加载单元 102 以随机顺序加载归一化谱，可以改善声音质量。当存储在归一化谱存储单元 101 中的归一化谱的数目较小时，这尤其有效。

[0115] 傅里叶逆变换单元 55 基于从分段选择单元 3 供应的分段以及从归一化谱加载单元 102 供应的归一化谱，生成音高波形，作为具有接近于音高周期的长度的话音波形（步骤 S2-3）。傅里叶逆变换单元 55 向音高波形叠加单元 56 输出所生成的音高波形。

[0116] 附带地，在该示例中，假设由分段选择单元 3 输出的浊音的分段（浊音分段）是幅度谱。因此，傅里叶逆变换单元 55 首先通过获取幅度谱和归一化谱的乘积来计算谱。随后，傅里叶逆变换单元 55 通过计算所计算的谱的傅里叶逆变换来生成音高波形（作为时域信号和话音波形）。

[0117] 音高波形叠加单元 56 通过在叠加由傅里叶逆变换单元 55 输出的多个音高波形时将其连接，而生成具有与由韵律生成单元 2 输出的韵律信息相一致或者相似的韵律的浊音波形（步骤 S2-4）。例如，音高波形叠加单元 56 通过采用以下参考文献 8 中描述的方法将音高波形叠加并且生成波形：

[0118] 参考文献 8：Eric Moulines, Francis Charpentier, ″ Pitch-synchronous Waveform Processing Techniques for Text-to-speech Synthesis Using Diphones, ″ (Netherlands), Elsevier Science Publishers B.V., Speech Communication, 第 9 卷，1990，第 453-467 页

[0119] 波形连接单元 7 通过连接由音高波形叠加单元 56 生成的浊音波形和由清音生成单元 6 生成的清音波形输出合成话音的波形（步骤 S2-5）。

[0120] 具体地，假设 v(t)(t = 1,2,3, …, t_v) 表示由音高波形叠加单元 56 生成的浊音波形，而 u(t)(t = 1,2,3, …, t_u) 表示由清音生成单元 6 生成的清音波形，波形连接单元 7 可以例如通过将浊音波形 v(t) 和清音波形 u(t) 连接来生成和输出以下合成话音波形 x(t)：

[0121] x(t) = v(t) 当 t = 1, …, t_v 时

[0122] x(t) = u(t-t_v) 当 t = (t_v+1), …, (t_v+t_u) 时

[0123] 在该示例性实施方式中，通过使用先前已经计算并且存储在归一化谱存储单元 101 中的归一化谱来生成并且输出合成话音的波形。因此，在生成合成话音时可以省略归一化谱的计算。从而，可以减少话音合成时必需的计算的数目。

[0124] 另外，由于归一化谱用于生成合成话音波形，所以与如在专利文献 1 中描述的设备中话音分段波形的周期分量和非周期分量用于生成合成话音的情况相比，可以生成更高声音质量的合成话音。

[0125] 〈第二示例性实施方式〉

[0126] 以下将参考附图描述根据本发明的话音合成器的第二示例性实施方式。该示例性实施方式的话音合成器通过与在第一示例性实施方式中采用的方法不同的方法来生成合成话音。图 6 是示出根据本发明的第二示例性实施方式的话音合成器的配置的示例的框图。

[0127] 如图 6 中所示，根据本发明的第二示例性实施方式的话音合成器包括傅里叶逆变换单元 91，代替图 1 中所示的第一示例性实施方式中的傅里叶逆变换单元 55。该示例性实施方式的话音合成器包括激励信号生成单元 92 和声道发音均衡滤波器 93，代替音高波形叠加单元 56。波形生成单元 4 不连接至分段选择单元 3 而是连接至分段选择单元 32。连接至分段选择单元 32 的是分段信息存储单元 122。其他组件与图 1 中所示的第一示例性实施方式中的话音合成器是等同的，并且由此为了简洁而省略了对其的重复说明，并且为其分配了与图 1 中相同的参考标记。

[0128] 分段信息存储单元 122 已经存储了线性预测分析参数（一种类型的声道发音均衡滤波器系数）作为分段信息。

[0129] 傅里叶逆变换单元 91 通过计算由归一化谱加载单元 102 输出的归一化谱的傅里叶逆变换来生成时域波形。傅里叶逆变换单元 91 向激励信号生成单元 92 输出所生成的时域波形。与图 1 中所示的第一示例性实施方式中的傅里叶逆变换单元 55 不同，傅里叶逆变换单元 91 的傅里叶逆变换计算的计算目标是归一化谱。由傅里叶逆变换单元 91 所采用的计算方法以及由傅里叶逆变换单元 91 输出的波形的长度与傅里叶逆变换单元 55 的等同。

[0130] 激励信号生成单元 92 通过在叠加由傅里叶逆变换单元 91 输出的多个时域波形时将其连接，而生成具有与由韵律生成单元 2 输出的韵律信息相一致或者相似的韵律的激励信号。激励信号生成单元 92 向声道发音均衡滤波器 93 输出所生成的激励信号。附带地，激励信号生成单元 92 通过在参考文献 8 中描述的方法（例如，类似于图 1 中所示的音高波形叠加单元 56）来将时域波形叠加并且生成波形。

[0131] 声道发音均衡滤波器 93 通过使用所选择的分段（由分段选择单元 32 输出）的声道发音均衡滤波器系数作为其滤波器系数，并且使用激励信号（由激励信号生成单元 92 输出）作为其滤波器输入信号，来向波形连接单元 7 输出浊音波形。在线性预测分析参数用作滤波器系数的情况下，声道发音均衡滤波器充当线性预测滤波器的反向滤波器，如以下参考文献 9 中所述：

[0132] 参考文献 9：Takashi Yahagi, " Digital Signal Processing and Basic Theories, " Corona Publishing Co., Ltd., 1996, 第 85-100 页

[0133] 波形连接单元 7 通过执行与第一示例性实施方式中的过程等同的过程来生成并且输出合成话音波形。

[0134] 以下将参考附图说明根据第二示例性实施方式的话音合成器的波形生成单元 4 的操作。图 7 是示出第二示例性实施方式中的话音合成器的波形生成单元 4 的操作的流程图。

[0135] 归一化谱加载单元 102 加载存储在归一化谱存储单元 101 中的归一化谱（步骤 S3-1）。随后，归一化谱加载单元 102 向傅里叶逆变换单元 91 输出加载的归一化谱（步骤 S3-2）。

[0136] 傅里叶逆变换单元 91 通过计算由归一化谱加载单元 102 输出的归一化谱的傅里叶逆变换来生成时域波形（步骤 S3-3）。傅里叶逆变换单元 91 向激励信号生成单元 92 输出所生成的时域波形。

[0137] 激励信号生成单元 92 基于由傅里叶逆变换单元 91 输出的多个时域波形来生成激励信号（步骤 S3-4）。

[0138] 声道发音均衡滤波器 93 通过使用来自分段选择单元 32 的所选择的分段的声道发音均衡滤波器系数作为其滤波器系数，并且使用来自激励信号生成单元 92 的激励信号作为其滤波器输入信号，来向波形连接单元 7 输出浊音波形（步骤 S3-5）。

[0139] 波形连接单元 7 通过执行与在第一示例性实施方式中的过程等同的过程来生成并且输出合成话音波形（步骤 S3-6）。

[0140] 该示例性实施方式的话音合成器基于归一化谱来生成激励信号，并且继而基于由激励信号通过声道发音均衡滤波器 93 的通过（滤波）而获得的浊音波形来生成合成话音波形。简言之，话音合成器通过与第一示例性实施方式的话音合成器采用的方法不同的方法来生成合成话音。

[0141] 根据该示例性实施方式，可以类似于第一示例性实施方式减少话音合成时必需的计算的数目。由此，即使在通过与由第一示例性实施方式中的话音合成器采用的方法不同的方法生成合成话音时，也有可能类似于第一示例性实施方式减少话音合成时必需的计算的数目。

[0142] 另外，由于类似于第一示例性实施方式，归一化谱用于生成合成话音波形，所以与如在专利文献 1 中描述的设备中话音分段波形的周期分量和非周期分量用于生成合成话音的情况相比，可以生成更高声音质量的合成话音。

[0143] 图 8 是示出根据本发明的话音合成器的主体部分的框图。如图 8 中所示，话音合成器 200 包括浊音生成单元 201（与图 1 或者图 6 中所示的浊音生成单元 5 相对应）、清音生成单元 202（与图 1 或者图 6 中所示的清音生成单元 6 相对应）以及合成话音生成单元 203（与图 1 或者图 6 中所示的波形连接单元 7 相对应）。浊音生成单元 201 包括归一化谱存储单元 204（与图 1 或者图 6 中所示的归一化谱存储单元 101 相对应）。

[0144] 归一化谱存储单元 204 预存储基于随机数序列计算的一个或多个归一化谱。浊音生成单元 201 基于与输入文本相对应的浊音的多个分段以及存储在归一化谱存储单元 204 中的归一化谱来生成浊音波形。

[0145] 清音生成单元 202 基于与文本相对应的清音的多个分段来生成清音波形。合成话音生成单元 203 基于由浊音生成单元 201 生成的浊音波形和由清音生成单元 202 生成的清音波形来生成合成话音。

[0146] 利用这样的配置，通过使用预存储在归一化谱存储单元 204 中的归一化谱生成合成话音的波形。由此，在生成合成话音时可以省略归一化谱的计算。从而，可以减少话音合成时必需的计算的数目。

[0147] 另外，由于话音合成器使用归一化谱来生成合成话音波形，所以与话音分段波形

的周期分量和非周期分量用于生成合成话音的情况相比，可以生成更高声音质量的合成话音。

[0148] 以上示例性实施方式中还公开了以下话音合成器 (1)-(5)：

[0149] (1) 话音合成器，其中，浊音生成单元 201 基于存储在归一化谱存储单元 204 中的归一化谱以及幅度谱来生成多个音高波形作为与文本相对应的浊音的分段，并且基于所生成的音高波形来生成浊音波形。

[0150] (2) 话音合成器，其中，浊音生成单元 201 基于存储在归一化谱存储单元 204 中的归一化谱来生成时域波形，基于所生成的时域波形和与输入文本相对应的韵律来生成激励信号，并且基于所生成的激励信号来生成浊音波形。

[0151] (3) 话音合成器，其中，通过使用基于随机数序列的群延迟来计算的一个或多个归一化谱预存储在归一化谱存储单元 204 中。

[0152] (4) 话音合成器，其中，归一化谱存储单元 204 预存储两个或者更多个归一化谱。浊音生成单元 201 通过使用与用于生成先前的浊音波形的归一化谱不同的归一化谱来生成每个浊音波形。利用这样的配置，可以防止由于归一化谱的不变相位分量而造成的合成话音的声音质量的退化。

[0153] (5) 话音合成器，其中，在归一化谱存储单元 204 中存储的归一化谱的数目在 2 至一百万的范围内。

[0154] 虽然以上已经参考示例性实施方式和示例描述了本发明，但是本发明不限于特定示出的示例性实施方式和示例。在本发明的范围内，可以对本发明的配置和细节做出本领域技术人员可理解的多种修改。

[0155] 本申请要求于 2010 年 3 月 25 日提交的日本专利申请号 2010-070378 的优先权，在此通过引用并入其全部公开内容。

[0156] 工业可应用性

[0157] 本发明可应用于多种生成合成话音的设备中。

[0158] 参考标记列表

[0159] 1 语言处理单元

[0160] 2 韵律生成单元

[0161] 3、32 分段选择单元

[0162] 4 波形生成单元

[0163] 5 浊音生成单元

[0164] 6 清音生成单元

[0165] 7 波形连接单元

[0166] 12、122 分段信息存储单元

[0167] 55、91 傅里叶逆变换单元

[0168] 56 音高波形叠加单元

[0169] 92 激励信号生成单元

[0170] 93 声道发音均衡滤波器

[0171] 101 归一化谱存储单元

[0172] 102 归一化谱加载单元

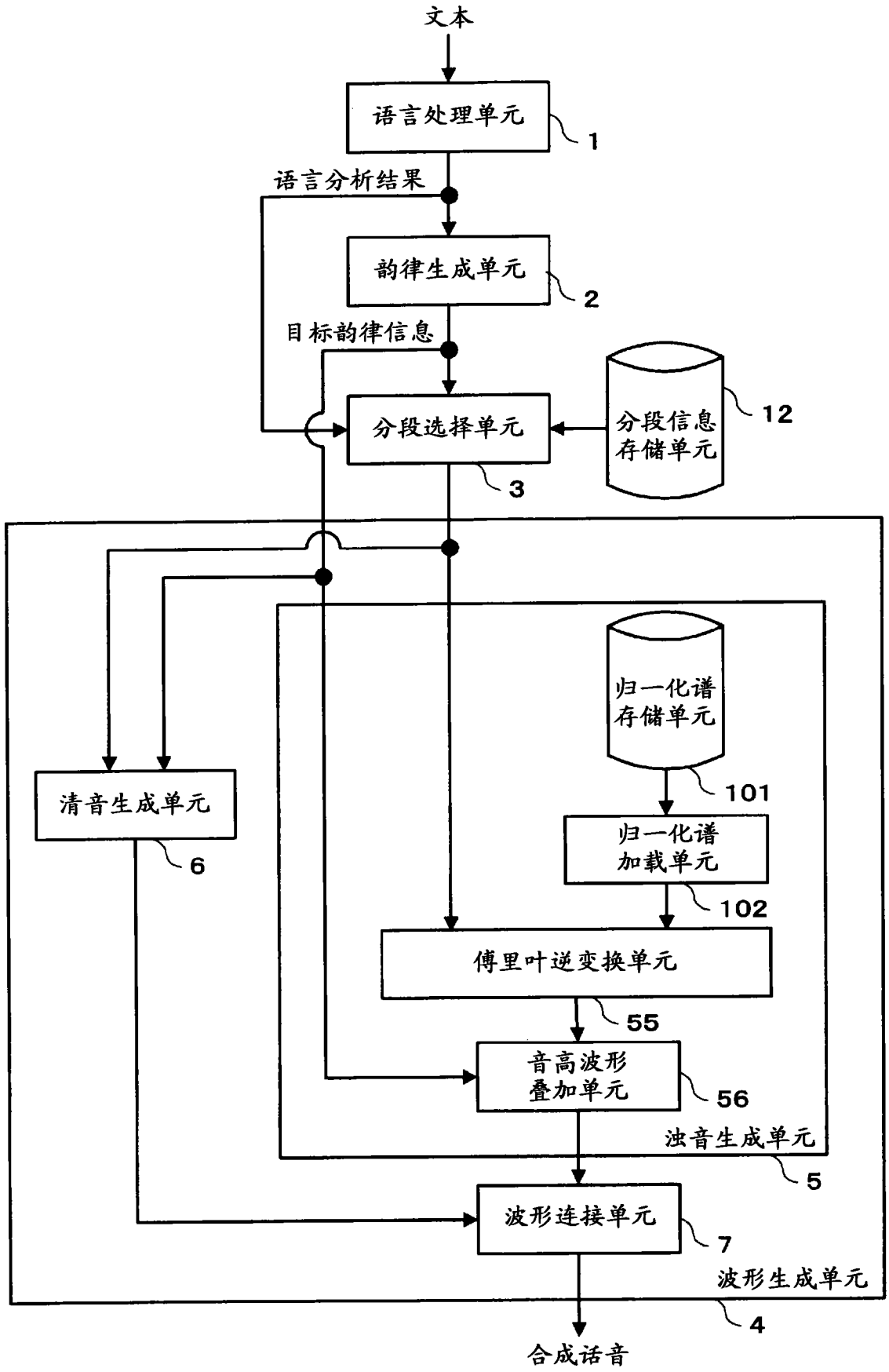

文本
语言处理单元
1
语言分析结果
韵律生成单元
2
目标韵律信息
分段选择单元
3
分段信息存储单元
12
归一化谱存储单元
101
归一化谱加载单元
102
清音生成单元
6
傅里叶逆变换单元
55
音高波形叠加单元
56
浊音生成单元
5
波形连接单元
7
波形生成单元
4
合成话音

图 1

	音高频率	持续时间	功率	与口音调核的距离
目标分段	pitch0	dur0	pow0	pos0
候选分段 A1	pitch1	dur1	pow1	pos1
候选分段 A2	pitch2	dur2	pow2	pos2

图 2

	开始边音高频率	结束边音高频率	开始边功率	结束边功率
候选分段 A1	pitch__beg1	pitch__end1	pow__beg1	pow__end1
候选分段 A2	pitch__beg2	pitch__end2	pow__beg2	pow__end2
候选分段 B1	pitch__beg3	pitch__end3	pow__beg3	pow__end3
候选分段 B2	pitch__beg4	pitch__end4	pow__beg4	pow__end4

图 3

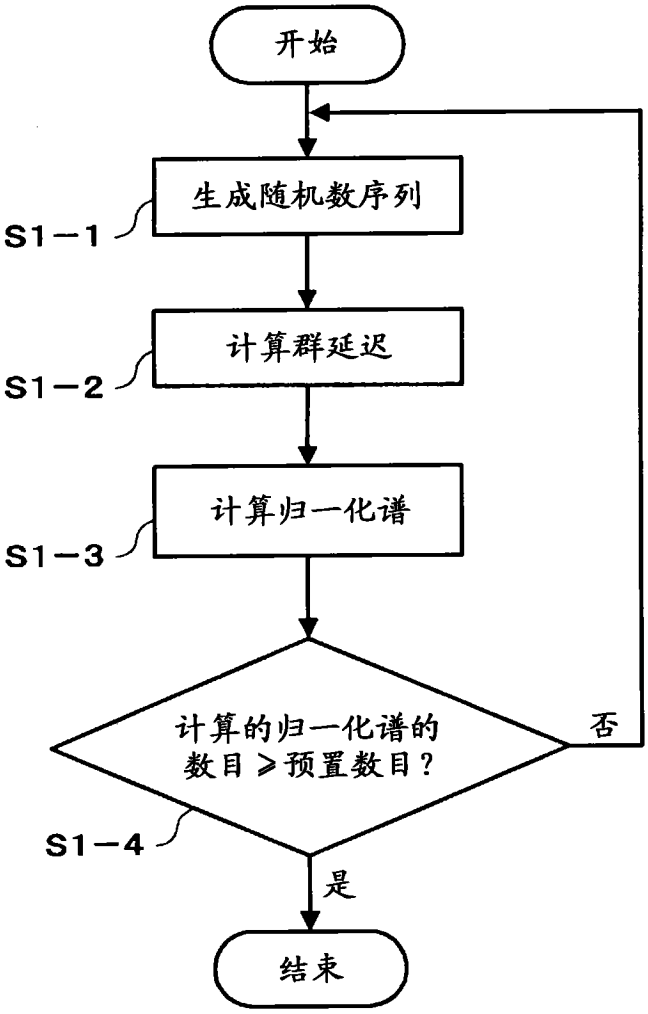

开始
生成随机数序列
S1－1
计算群延迟
S1－2
计算归一化谱
S1－3
计算的归一化谱的
数目≥预置数目？
否
S1－4
是
结束

图 4

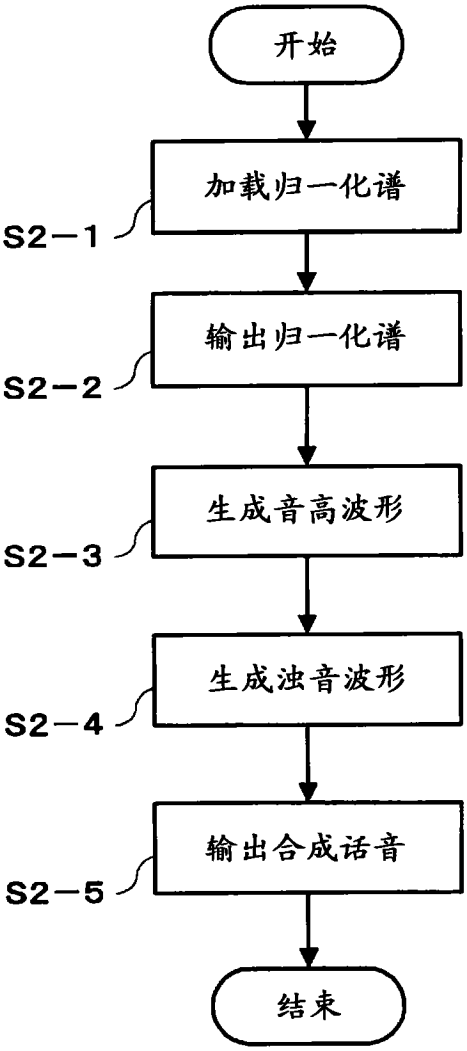

开始
加载归一化谱
S2－1
输出归一化谱
S2－2
生成音高波形
S2－3
生成浊音波形
S2－4
输出合成话音
S2－5
结束

图 5

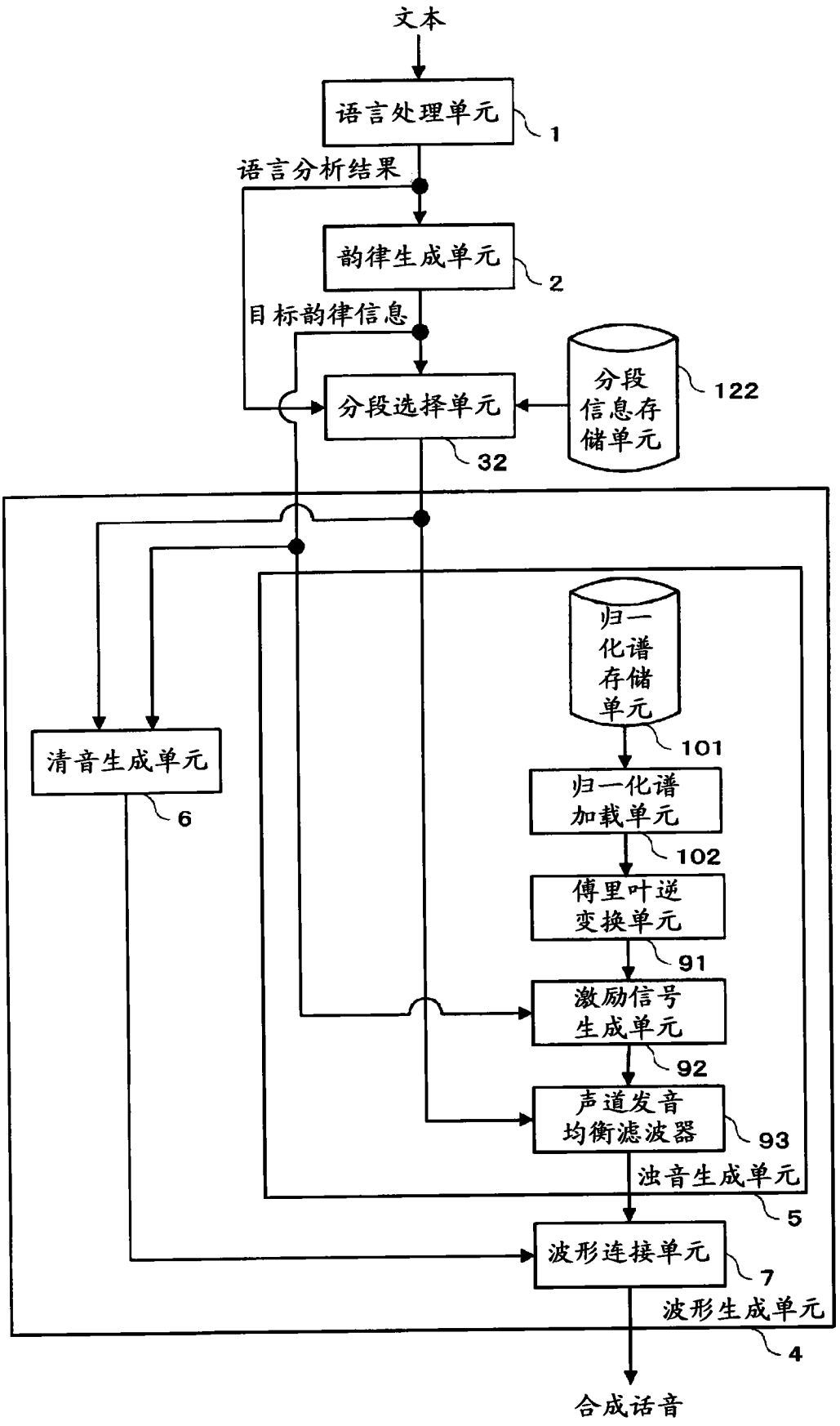

文本
语言处理单元
1
语言分析结果
韵律生成单元
2
目标韵律信息
分段信息存储单元
122
分段选择单元
32
归一化谱存储单元
101
清音生成单元
6
归一化谱加载单元
102
傅里叶逆变换单元
91
激励信号生成单元
92
声道发音均衡滤波器
93
浊音生成单元
5
波形连接单元
7
波形生成单元
4
合成话音

图 6

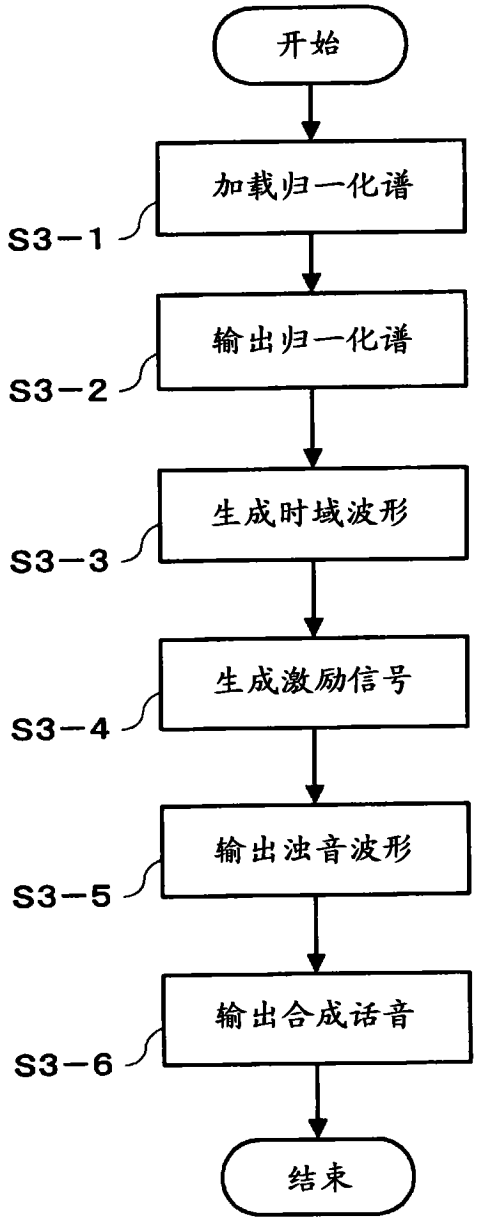

开始
加载归一化谱
S3-1
输出归一化谱
S3-2
生成时域波形
S3-3
生成激励信号
S3-4
输出浊音波形
S3-5
输出合成话音
S3-6
结束

图 7

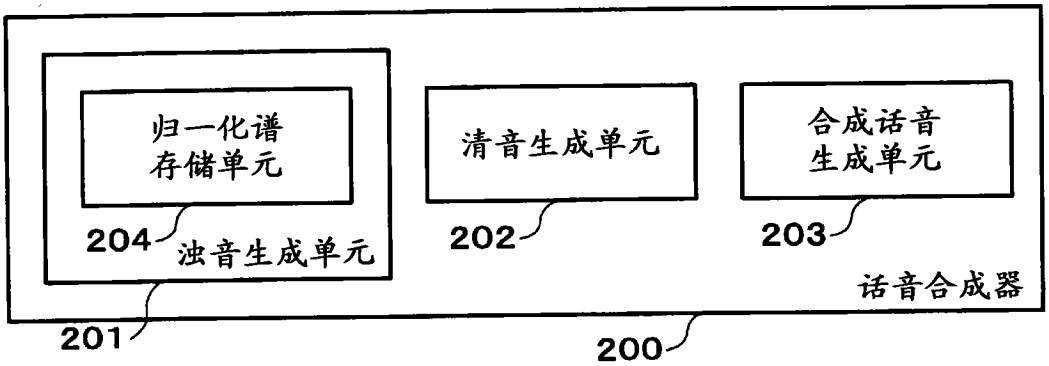

归一化谱
存储单元
204
浊音生成单元
201
清音生成单元
202
合成话音
生成单元
203
话音合成器
200

图 8