[19] 中华人民共和国国家知识产权局

[51] Int. Cl.
G10L 21/04 (2006.01)
G10L 13/06 (2006.01)

[12] 发明专利申请公布说明书

[21] 申请号 200880001672.7

[43] 公开日 2009 年 11 月 11 日

[11] 公开号 CN 101578659A

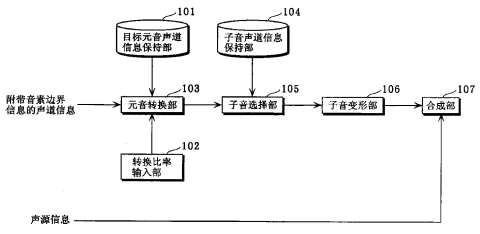
[22] 申请日 2008.5.8
[21] 申请号 200880001672.7
[30] 优先权
[32] 2007. 5.14 [33] JP [31] 128555/2007
[86] 国际申请 PCT/JP2008/001160 2008.5.8
[87] 国际公布 WO2008/142836 日 2008.11.27
[85] 进入国家阶段日期 2009.7.3
[71] 申请人 松下电器产业株式会社
地址 日本大阪府
[72] 发明人 广濑良文 釜井孝浩 加藤弓子
[74] 专利代理机构 永新专利商标代理有限公司
代理人 杨 谦 胡建新
权利要求书 7 页 说明书 38 页 附图 23 页

[54] 发明名称

音质转换装置及音质转换方法

[57] 摘要

一种音质转换装置，利用与输入声音对应的信息来转换输入声音的音质，包括：目标元音声道信息保持部(101)，按每个元音来保持目标元音声道信息，所述目标元音声道信息是表示成为目标的音质的元音的声道信息；元音转换部(103)，接受被付与了输入声音所对应的音素及音素的时间长度信息的声道信息、即附带音素边界信息的声道信息，将所述附带音素边界信息的声道信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的所述目标元音声道信息保持部(101)所保持的声道信息的时间变化以第二函数进行近似，通过结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息；以及合成部(107)，利用由所述元音转换部(103)转换后的元音的声道信息，合成声音。



1、一种音质转换装置，利用与输入声音对应的信息来转换输入声音的音质，包括：

目标元音声道信息保持部，按每个元音来保持目标元音声道信息，所述目标元音声道信息是表示成为目标的音质的元音的声道信息；

元音转换部，接受被付与了输入声音所对应的音素及音素的时间长度信息的声道信息、即附带音素边界信息的声道信息，将所述附带音素边界信息的声道信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的所述目标元音声道信息保持部所保持的声道信息的时间变化以第二函数进行近似，通过结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息；以及

合成部，利用由所述元音转换部转换后的元音的声道信息，合成声音。

2、如权利要求 1 所述的音质转换装置，还包括：

子音声道信息导出部，接受所述附带音素边界信息的声道信息，并按该附带音素边界信息的声道信息所包含的每个子音的声道信息，从包含所述成为目标的音质以外的音质的子音的声道信息之中，导出具有与所述附带音素边界信息的声道信息所包含的子音相同的音素的子音的声道信息，

所述合成部利用由所述元音转换部转换后的元音的声道信息，和在所述子音声道信息导出部导出的子音的声道信息，合成声音。

3、如权利要求 2 所述的音质转换装置，

所述子音声道信息导出部具有：

子音声道信息保持部，按每个子音保持从多个讲话者的声音抽取了的声道信息；以及

子音选择部，接受所述附带音素边界信息的声道信息，并按该附带音素边界信息的声道信息所包含的每个子音的声道信息，将适合于位于该子音之前或之后的元音区间的由所述元音转换部转换后的元音的声道信息的、具有与该子音相同的音素的子音的声道信息，从所述子音声道信息保持部所保持的子音的声道信息中选择。

4、如权利要求 3 所述的音质转换装置，

所述子音选择部，接受所述附带音素边界信息的声道信息，依据该附带音素边界信息的声道信息所包含的每个子音的声道信息的数值与位于该子音之前或之后的元音区间的、由所述元音转换部转换后的元音的声道信息的数值的连续性，从所述子音声道信息保持部所保持的子音的声道信息中选择具有与该子音相同的音素的子音的声道信息。

5、如权利要求 3 所述的音质转换装置，

还包括子音变形部，将在所述子音选择部选择的子音的声道信息进行变形，以使该子音的声道信息的数值与位于该子音之后的元音区间的、由所述元音转换部转换后的元音的声道信息的数值的连续性变好。

6、如权利要求 1 所述的音质转换装置，

还包括转换比率输入部，输入表示向成为目标的音质转换的程度的转换比率，

所述元音转换部接受被付与了输入声音所对应的音素及音素的时间长度信息的声道信息、即附带音素边界信息的声道信息和在所述转换比率输入部输入的所述转换比率，将所述附带音素边界信息的声道

信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的所述目标元音声道信息保持部所保持的声道信息的时间变化以第二函数进行近似，通过以所述转换比率结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息。

7、如权利要求 6 所述的音质转换装置，

所述元音转换部，以次数为单位将所述附带音素边界信息的声道信息所包含的元音的声道信息以第一多项式进行近似，以次数为单位将与该元音相同的元音的所述目标元音声道信息保持部所保持的所述目标元音声道信息以第二多项式进行近似，并以次数为单位，通过以所述转换比率混合所述第一多项式的系数和所述第二多项式的系数，从而求出第三多项式的各个次数的系数，将转换后的元音的声道信息以所述第三多项式进行近似。

8、如权利要求 1 所述的音质转换装置，

所述元音转换部进一步，将包含作为第一元音的声道信息和第二元音的声道信息之间在时间上的边界的元音边界的规定时间设为过渡区间，对该过渡区间中所包含的所述第一元音的声道信息和所述第二元音的声道信息进行插值，以使在所述元音边界中所述第一元音的声道信息和所述第二元音的声道信息被连续地连接。

9、如权利要求 8 所述的音质转换装置，

所述规定时间被设定为，位于所述元音边界的前后的、所述第一元音和所述第二元音的持续时间长度越长，所述规定时间就越长。

10、如权利要求 1 所述的音质转换装置，

所述声道信息是 PARCOR 系数或者声道声管模型的反射系数。

11、如权利要求 10 所述的音质转换装置，

所述 PARCOR 系数或者声道声管模型的反射系数通过对输入声音进行 LPC 分析，并根据分析出的全极点模型的多项式被算出。

12、如权利要求 10 所述的音质转换装置，

所述 PARCOR 系数或者声道声管模型的反射系数通过对输入声音进行 ARX 分析，并根据分析后的全极点模型的多项式被算出。

13、如权利要求 1 所述的音质转换装置，

所述附带音素边界信息的声道信息根据从文本生成的合成声音来决定。

14、如权利要求 1 所述的音质转换装置，

所述目标元音声道信息保持部，保持根据稳定元音区间抽取部和目标声道信息制作部制作的目标元音声道信息，

所述稳定元音区间抽取部从成为目标的音质的声音，检测稳定的元音区间，

所述目标声道信息制作部从稳定的元音区间，抽取成为目标的声道信息。

15、如权利要求 14 所述的音质转换装置，

所述稳定元音区间抽取部具有：

音素识别部，识别成为所述目标的音质的声音中所包含的音素；以及

稳定区间抽取部，在所述音素识别部识别的元音区间中，将所述音素识别部中的识别结果的似然比规定的阈值高的区间作为稳定元音区间进行抽取。

16、一种音质转换方法，用于利用与输入声音对应的信息来转换输入声音的音质，包括：

元音转换步骤，接受被付与了输入声音所对应的音素及音素的时间长度信息的声道信息、即附带音素边界信息的声道信息，将所述附带音素边界信息的声道信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的所述目标元音声道信息保持部所保持的声道信息的时间变化以第二函数进行近似，通过结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息；以及

合成步骤，利用由所述元音转换步骤转换后的元音的声道信息，合成声音。

17、一种程序，用于利用与输入声音对应的信息来转换输入声音的音质，使计算机执行以下步骤：

元音转换步骤，接受被付与了输入声音所对应的音素及音素的时间长度信息的声道信息、即附带音素边界信息的声道信息，将所述附带音素边界信息的声道信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的所述目标元音声道信息保持部所保持的声道信息的时间变化以第二函数进行近似，通过结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息；以及

合成步骤，利用由所述元音转换步骤转换后的元音的声道信息，合成声音。

18、一种音质转换系统，用于利用与被转换声音对应的信息来转换被转换声音的音质，所述音质转换系统包括：

服务器；以及

终端，与所述服务器通过网络相连接，

所述服务器包括：

目标元音声道信息保持部，按每个元音来保持目标元音声道信息，所述目标元音声道信息是表示成为目标的音质的元音的声道信息；

目标元音声道信息发送部，将所述目标元音声道信息保持部所保持的目标元音声道信息通过网络发送到所述终端；

被转换声音保持部，保持作为被转换声音所对应的信息的被转换声音信息；以及

被转换声音信息发送部，将所述被转换声音保持部所保持的被转换声音信息通过网络发送到所述终端，

所述终端包括：

目标元音声道信息接收部，接收由所述目标元音声道信息发送部发送了的所述目标元音声道信息；

被转换声音信息接收部，接收由所述被转换声音信息发送部发送了的所述被转换声音信息；

元音转换部，将由所述被转换声音信息接收部接收了的被转换声音信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的由所述目标元音声道信息接收部接收了的所述目标元音声道信息的时间变化以第二函数进行近似，通过结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息；以及

合成部，利用根据所述元音转换部的转换后的元音的声道信息，合成声音。

19、一种音质转换系统，用于利用与被转换声音对应的信息来转换被转换声音的音质，所述音质转换系统包括：

终端；以及

服务器，与所述终端通过网络相连接，

所述终端包括：

目标元音声道信息制作部，按每个元音，保持并制作目标元音声道信息，所述目标元音声道信息是表示成为目标的音质的元音的声道信息；

目标元音声道信息发送部，将所述目标元音声道信息制作部所制作的所述目标元音声道信息通过网络发送到所述终端；

音质转换声音接收部，从所述服务器接收音质转换后的声音；以及

再生部，再生所述音质转换声音接收部接收了的所述音质转换后的声音，

所述服务器包括：

被转换声音保持部，保持作为被转换声音所对应的信息的被转换声音信息；

目标元音声道信息接收部，接收由所述目标元音声道信息发送部发送了的所述目标元音声道信息；

元音转换部，将所述被转换声音保持部所保持的被转换声音信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的由所述目标元音声道信息接收部接收了的所述目标元音声道信息的时间变化以第二函数进行近似，通过结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息；

合成部，利用根据所述元音转换部的转换后的元音的声道信息，合成声音；以及

合成声音发送部，将在合成部中合成后的声音作为音质转换后的声音，通过网络发送到所述音质转换声音接收部。

音质转换装置及音质转换方法

技术领域

本发明涉及转换声音的音质的音质转换装置及音质转换方法，尤其涉及将输入声音的音质转换为作为目标的讲话者的声音的音质的音质转换装置及音质转换方法。

背景技术

近几年，随着声音合成技术的发展，已经能够制作出极高音质的合成音。

但是，以往的合成音的用途主要以播音员的风格朗读新闻等用途为中心。

另一方面，在移动电话服务等领域，提供使用名人的声音信息来代替铃声之类的服务等，有特征的声音（个人再现性高的合成音，以及女高中生腔调或者关西方言腔调等具有特征性的韵律和音质的合成音）作为一个内容开始流通。为了增加这样的人际交流中的乐趣，可以想像对于制作给对方听的特征性的声音的要求今后会增高。

再者，作为合成音的方法，大致分为以下两种。即，从预先准备好的声音单元 DB(数据库)中选择适当的声音单元，并通过将其进行连接来合成音的波形连接型声音合成方法，和对声音进行分析，以分析后的参数为基础来合成声音的分析合成型声音合成方法。

如果考虑使上述合成音的音质进行各种各样的变化，则在波形连接型声音合成方法中，需要尽量准备必要的声音单元 DB，并切换声音单元 DB，同时需要对声音单元进行连接。因此，为了制作各种各样音质的合成音，需要庞大的费用。

另一方面，在分析合成型声音合成方法中，通过使分析后的声音参数进行变形，能够转换合成音的音质。作为参数的变形方法，存在使用作为同样的讲话内容的不同的两个讲话进行转换的方法。

专利文献 1 表示使用神经网络等学习模式的分析合成型声音合成方法的一个例子。

图 1 是表示利用专利文献 1 的付与感情方法的声音处理系统的构成的图。

此图所示的声音处理系统包括：声分析部 2、频谱的 DP(Dynamic Programming：动态编程)匹配部 4、各个音素的时间长度伸缩部 6、神经网络部 8、依据规则的合成参数生成部、时间长度伸缩部、声音合成系统部。声音处理系统在通过神经网络部 8 进行用于将无感情的声音的声特征参数转换成有感情的声音的声特征参数的学习之后，使用学习完毕的该神经网络部 8，将感情付与无感情的声音。

频谱的 DP 匹配部 4 对声分析部 2 所抽取的特征参数之中的、无感情的声音的频谱的特征参数和有感情的声音的频谱的特征参数之间的相似度进行每时每刻的调查，通过取得每个同样的音素的时间上的对应，从而求出针对无感情声音的感情声音的每个音素的时间上的伸缩率。

各个音素的时间长度伸缩部 6 按照频谱的 DP 匹配部 4 所取得的每个音素的时间上的伸缩率，将感情声音的特征参数的时间序列在时间上进行归一化，从而使其适合无感情声音。

神经网络部 8 在学习时，每时每刻都对给予输入层的无感情声音的声特征参数和给予输出层的感情声音的声特征参数的差别进行学习。

并且，神经网络部 8 在感情的付与时，利用在学习时所决定的网络内部的加权系数，每时每刻进行由给予输入层的无感情声音的声特征参数推算感情声音的声响特征参数的计算。如上所述的，是根据学习模式进行从无感情声音到感情声音的转换。

但是，专利文献 1 的技术需要记录预先决定了的与用于学习的文章同样的内容为目标的有感情的发音。因此，在将专利文献 1 的技术应用于转换讲话者的情况下，需要使作为目标的讲话者将预先决定了的用于学习的文章全部念出来。所以，存在对目标讲话者增加负担的问题。

作为不必将预先决定了的用于学习的文章念出来也可以的方法，具有专利文献 2 中记述的方法。专利文献 2 中记述的方法是通过文本合成装置合成同样的讲话内容，并根据合成后的声音与目标声音的差分，来编写声音频谱形状的变换函数的方法。

图 2 是专利文献 2 的音质转换装置的框图。

目标讲话者的声音信号被输入目标讲话者声音输入部 11a，声音识别部 19 对被输入到目标讲话者声音输入部 11a 的目标讲话者声音进行声音识别，将目标讲话者声音的发音内容与音标一起输出到音标序列输入部 12a。声音合成部 14 按照被输入的音标序列，利用声音合成用数据存储部 13 内的声音合成用数据库来制作合成音。目标讲话者声音特征参数抽取部 15 对目标讲话者声音进行分析从而抽取特征参数，合成音特征参数抽取部 16 对制作成的合成音进行分析从而抽取特征参数。变换函数生成部 17 利用抽取了的双方的特征参数，生成将合成音的频谱形状转换为目标讲话者声音的频谱形状的函数。音质转换部 18 根据生成了的变换函数，进行输入信号的音质转换。

如上所述，因为将目标讲话者声音的声音识别结果作为用于合成音生成的音标序列而输入声音合成部 14，所以用户不需要以文本等输入音标序列，从而能够谋求处理的自动化。

并且，作为以较少的存储量即能够生成多个音质的声音合成装置，存在专利文献 3 的声音合成装置。专利文献 3 所涉及的声音合成装置包含：声音单元存储部、多个元音单元存储部、多个基频存储部。声音单元存储部保持包含元音的过渡部分的子音单元。各个元音单元存储部存储讲话者一个人的元音单元。多个基频存储部分别存储成为元

音单元的基础的讲话者的基频。

声音合成装置从多个元音单元存储部中读出被指定的讲话者的元音单元，并通过与存储在声音单元存储部中的预先决定了的子音单元连接，来合成声音。因此，能够将输入声音的音质转换为被指定的讲话者的音质。

专利文献 1：（日本）特开平 7－72900 号公报(第 3－8 页，图 1)

专利文献 2：（日本）特开 2005－266349 号公报(第 9－10 页，图 2)

专利文献 3：（日本）特开平 5－257494 号公报

在专利文献 2 的技术中，通过声音识别部 19 识别目标讲话者的讲话内容，从而生成音标序列，利用保持在标准的声音合成用数据存储部 13 的数据，声音合成部 14 合成合成音。但是，普遍存在无法避免声音识别部 19 产生识别错误的问题。并且无法避免给在变换函数生成部 17 编写的变换函数的性能带来巨大的影响。而且，通过变换函数生成部 17 编写的变换函数是，从声音合成用数据存储部 13 所保持的声音的音质转换为目标讲话者的音质的变换函数。因此，存在通过音质转换部 18 转换的被转换输入信号，与声音合成用数据存储部 13 的音质相同，或者在不是极其相似的音质的声音信号的情况下，转换后输出信号不一定与目标讲话者的音质一致的问题。

而且，专利文献 3 所涉及的声音合成装置，通过切换目标元音的一帧的音质特征，来进行输入声音的音质转换。因此，只能够将输入声音的音质转换为预先登记了的讲话者的音质，而不能够生成介于多个讲话者的音质的中间的音质的声音。并且，由于仅使用一帧的音质特征来进行音质的转换，所以存在连续发音中的自然劣化大的问题。

进一步，在专利文献 3 所涉及的声音合成装置中，在通过元音单元的置换而使元音特征被大为转换的情况下，存在预先被唯一决定的子音特征和转换后的元音特征之间的差变大的情况。在此情况下，为

了使两者的差变小，即使在元音特征及子音特征之间进行了插值，也存在合成音的自然性大为劣化之类的问题。

发明内容

本发明就是为了解决上述以往的问题，其目的在于，提供一种能够实现对被转换输入信号没有限制的音质转换的音质转换装置及音质转换方法。

并且，本发明的目的在于，提供一种不受目标讲话者的讲话的识别错误的影响，就能够对被转换输入信号进行音质转换的音质转换装置及音质装换方法。

本发明的某个局面所涉及的音质转换装置，利用与输入声音对应的信息来转换输入声音的音质，包括：目标元音声道信息保持部，按每个元音来保持目标元音声道信息，所述目标元音声道信息是表示成为目标的音质的元音的声道信息；元音转换部，接受被付与了输入声音所对应的音素及音素的时间长度信息的声道信息、即附带音素边界信息的声道信息，将所述附带音素边界信息的声道信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的所述目标元音声道信息保持部所保持的声道信息的时间变化以第二函数进行近似，通过结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息；以及合成部，利用由所述元音转换部转换后的元音的声道信息，合成声音。

根据此构成，利用目标元音声道信息保持部所保持的目标元音声道信息，进行声道信息的转换。这样，因为能够将目标元音声道信息作为绝对的目标来利用，所以可以完全不限制转换前的声音的音质，从而可以输入任何音质的声音。即，因为对被输入的被转换声音的限制非常少，所以能够针对广泛的声音进行音质的转换。

最好是，所述音质转换装置还包括：子音声道信息导出部，接受所述附带音素边界信息的声道信息，并按该附带音素边界信息的声道

信息所包含的每个子音的声道信息，从包含所述成为目标的音质以外的音质的子音的声道信息之中，导出具有与所述附带音素边界信息的声道信息所包含的子音相同的音素的子音的声道信息，所述合成部利用由所述元音转换部转换后的元音的声道信息，和在所述子音声道信息导出部导出的子音的声道信息，合成声音。

进而最好是，所述子音声道信息导出部具有：子音声道信息保持部，按每个子音保持从多个讲话者的声音抽取了的声道信息；以及子音选择部，接受所述附带音素边界信息的声道信息，并按该附带音素边界信息的声道信息所包含的每个子音的声道信息，将适合于位于该子音之前或之后的元音区间的、由所述元音转换部转换后的元音的声道信息的、具有与该子音相同的音素的子音的声道信息，从所述子音声道信息保持部所保持的子音的声道信息中选择。

进而最好是，所述子音选择部，接受所述附带音素边界信息的声道信息，依据该附带音素边界信息的声道信息所包含的每个子音的声道信息的数值，与位于该子音之前或之后的元音区间的、由所述元音转换部转换后的元音的声道信息的数值的连续性，从所述子音声道信息保持部所保持的子音的声道信息中选择具有与该子音相同的音素的子音的声道信息。

因此，能够使用适合于转换后的元音的声道信息的最合适的子音声道信息。

进而最好是，所述音质转换装置还包括转换比率输入部，输入表示向成为目标的音质转换的程度的转换比率，所述元音转换部接受被付与了输入声音所对应的音素及音素的时间长度信息的声道信息、即附带音素边界信息的声道信息，和在所述转换比率输入部输入的所述转换比率，将所述附带音素边界信息的声道信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的所述目标元音声道信息保持部所保持的声道信息的时间变化以第二函数进行近似，通过以所述转换比率结合所述第一函数和所述第二函数从而

求出第三函数，并由所述第三函数生成转换后的元音的声道信息。

因此，能够控制成为目标的音质的强调程度。

进而最好是，所述目标元音声道信息保持部，保持根据稳定元音区间抽取部和目标声道信息制作部制作的目标元音声道信息，所述稳定元音区间抽取部从成为目标的音质的声音，检测稳定的元音区间，所述目标声道信息制作部从稳定的元音区间，抽取成为目标的声道信息。

而且，作为成为目标的音质的声道信息，只保持稳定的元音区间的声道信息即可。并且，在目标讲话者的讲话的识别时，只在元音稳定区间中进行音素识别即可。因此，不发生目标讲话者的讲话的识别错误。因而，不受目标讲话者的识别错误的影响，就能够对被转换输入信号进行音质转换。

本发明的其他局面所涉及的音质转换系统，用于利用与被转换声音对应的信息来转换被转换声音的音质，所述音质转换系统包括：服务器；以及终端，与所述服务器通过网络相连接。所述服务器包括：目标元音声道信息保持部，按每个元音来保持目标元音声道信息，所述目标元音声道信息是表示成为目标的音质的元音的声道信息；目标元音声道信息发送部，将所述目标元音声道信息保持部所保持的目标元音声道信息通过网络发送到所述终端；被转换声音保持部，保持作为被转换声音所对应的信息的被转换声音信息；以及被转换声音信息发送部，将所述被转换声音保持部所保持的被转换声音信息通过网络发送到所述终端。所述终端包括：目标元音声道信息接收部，接收由所述目标元音声道信息发送部发送了的所述目标元音声道信息；被转换声音信息接收部，接收由所述被转换声音信息发送部发送了的所述被转换声音信息；元音转换部，将由所述被转换声音信息接收部接收了的被转换声音信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的由所述目标元音声道信息接收部接收了的所述目标元音声道信息的时间变化以第二函数进行近似，通

过结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息；以及合成部，利用由所述元音转换部转换后的元音的声道信息，合成声音。

利用终端的用户下载被转换声音信息和元音目标声道信息，并能够在终端进行被转换声音信息的音质转换。例如，在被转换声音信息是声音内容的情况下，用户能够以适合自己的爱好的音质来再生声音内容。

本发明的另一个其他局面所涉及的音质转换系统，用于利用与被转换声音对应的信息来转换被转换声音的音质，所述音质转换系统包括：终端；以及服务器，与所述终端通过网络相连接。所述终端包括：目标元音声道信息制作部，按每个元音，保持并制作目标元音声道信息，所述目标元音声道信息是表示成为目标的音质的元音的声道信息；目标元音声道信息发送部，将所述目标元音声道信息制作部所制作的所述目标元音声道信息通过网络发送到所述终端；音质转换声音接收部，从所述服务器接收音质转换后的声音；以及再生部，再生所述音质转换声音接收部接收了的所述音质转换后的声音。所述服务器包括：被转换声音保持部，保持作为被转换声音所对应的信息的被转换声音信息；目标元音声道信息接收部，接收由所述目标元音声道信息发送部发送了的所述目标元音声道信息；元音转换部，将所述被转换声音保持部所保持的被转换声音信息所包含的元音的声道信息的时间变化以第一函数进行近似，将与该元音相同的元音的由所述目标元音声道信息接收部接收了的所述目标元音声道信息的时间变化以第二函数进行近似，通过结合所述第一函数和所述第二函数从而求出第三函数，并由所述第三函数生成转换后的元音的声道信息；合成部，利用由所述元音转换部转换后的元音的声道信息，合成声音；以及合成声音发送部，将在合成部中合成后的声音作为音质转换后的声音，通过网络发送到所述音质转换声音接收部。

终端制作并发送目标元音声道信息，且接收并再生通过服务器转

换了音质的声音。因此，只需在终端制作成为目标的元音的声道信息即可，从而能够大为减小处理负荷。而且，终端的用户能够以适合自己的爱好的音质来收听适合自己的爱好的声音内容。

并且，本发明不仅可以作为具备如此特征性单元的音质转换装置来实现，还可以作为将音质转换装置所包括的特征性单元作为步骤的音质转换方法来实现，或将音质转换方法中所包括的特征性步骤作为使计算机执行的程序来实现。并且，不用说，能够使这样的程序通过CD－ROM(Compact Disc－Read Only Memory：只读存储光盘)等的记录介质或互联网等的通信网络流通。

根据本发明，作为目标讲话者的信息，只准备元音稳定区间的信息即可，能够大幅度减少对目标讲话者的负担。例如，在日语的情况下，只需准备 5 个元音即可。因而，能够容易进行音质转换。

而且，因为作为目标讲话者的信息，只需识别元音稳定区间的声道信息即可，所以不需要象专利文献 2 的以往技术那样，对整个目标讲话者的发音进行识别，从而由于声音识别错误的影响较小。

并且，在专利文献 2 的现有技术中，因为根据声音合成部的声音单元与目标讲话者的发音的差分编写变换函数，所以需要被转换声音的音质与声音合成部所保持的声音单元的音质相同或相似，而本发明的音质转换装置将目标讲话者的元音声道信息作为绝对的目标。因此，不限定转换前的声音的音质，输入任何音质的声音都可以。即，因为对被输入的被转换声音的限制非常少，所以能够对广泛的声音进行音质转换。

而且，有关目标讲话者的信息，因为只要保持元音稳定区间的信息即可，所以只需很小的存储容量即可，因此，能够在通过便携终端或网络的服务等上利用。

附图说明

图 1 是表示以往的声音处理系统的构成的图。

图 2 是表示以往的音质转换装置的构成的图。

图 3 是表示本发明的实施例 1 所涉及的音质转换装置的构成的图。

图 4 是表示声道截面面积函数和 PARCOR 系数（偏自相关系数）的关系的图。

图 5 是表示生成目标元音声道信息保持部所保持的目标元音声道信息的处理部的构成的图。

图 6 是表示生成目标元音声道信息保持部所保持的目标元音声道信息的处理部的构成的图。

图 7 是表示元音的稳定区间的一个例子的图。

图 8A 是表示被输入的附带音素边界信息的声道信息的制作方法的一个例子的图。

图 8B 是表示被输入的附带音素边界信息的声道信息的制作方法的一个例子的图。

图 9 是表示利用了文本声音合成装置的、被输入的附带音素边界信息的声道信息的制作方法的一个例子的图。

图 10A 是表示根据元音/a/的一次 PARCOR 系数的声道信息的一个例子的图。

图 10B 是表示根据元音/a/的二次 PARCOR 系数的声道信息的一个例子的图。

图 10C 是表示根据元音/a/的三次 PARCOR 系数的声道信息的一个例子的图。

图 10D 是表示根据元音/a/的四次 PARCOR 系数的声道信息的一个例子的图。

图 10E 是表示根据元音/a/的五次 PARCOR 系数的声道信息的一个例子的图。

图 10F 是表示根据元音/a/的六次 PARCOR 系数的声道信息的一个例子的图。

图 10G 是表示根据元音/a/的七次 PARCOR 系数的声道信息的一

个例子的图。

图 10H 是表示根据元音/a/的八次 PARCOR 系数的声道信息的一个例子的图。

图 10I 是表示根据元音/a/的九次 PARCOR 系数的声道信息的一个例子的图。

图 10J 是表示根据元音/a/的十次 PARCOR 系数的声道信息的一个例子的图。

图 11A 是表示根据元音转换部的元音的声道形状的多项式近似的具体例子的图。

图 11B 是表示根据元音转换部的元音的声道形状的多项式近似的具体例子的图。

图 11C 是表示根据元音转换部的元音的声道形状的多项式近似的具体例子的图。

图 11D 是表示根据元音转换部的元音的声道形状的多项式近似的具体例子的图。

图 12 是表示根据元音转换部的元音区间的 PARCOR 系数被转换的情况的图。

图 13 是对关于设置过渡区间，以对 PARCOR 系数的值进行插值的例子进行说明的图。

图 14A 是表示在对元音/a/和元音/i/的边界的 PARCOR 系数进行插值的情况下的频谱的图。

图 14B 是表示将元音/a/和元音/i/的边界的声音通过平滑转换进行连接的情况下的频谱的图。

图 15 是从对合成后的 PARCOR 系数进行插值后的 PARCOR 系数中再次抽取共振峰并绘制的图形。

图 16 是，在图 16(a)为/a/和/u/的连接，图 16(b)为/a/和/e/的连接，图 16(c)为/a/和/o/的连接之时的，表示根据平滑转换连接的频谱、对 PARCOR 系数进行插值后的频谱以及根据 PARCOR 系数插值

的共振峰的移动的图。

图 17A 是表示转换前的男性讲话者的声道截面面积的情况的图。

图 17B 是表示目标讲话者的女性的声道截面面积的情况的图。

图 17C 是表示与以转换比率 50%对转换前的 PARCOR 系数进行转换后的 PARCOR 系数相对应的声道截面面积的情况的图。

图 18 是用于说明通过子音选择部选择子音声道信息的处理的模式图。

图 19A 是目标元音声道信息保持部的构筑处理的流程图。

图 19B 是将被输入了的附带音素边界信息的声音转换为目标讲话者的声音的处理的流程图。

图 20 是表示本发明的实施例 2 所涉及的音质转换系统的构成的图。

图 21 是表示本发明的实施例 2 所涉及的音质转换系统的工作的图。

图 22 是表示本发明的实施例 3 所涉及的音质转换系统的构成的图。

图 23 是表示本发明的实施例 3 所涉及的音质转换系统的处理的流程的流程图。

附图标记说明

101 目标元音声道信息保持部

102 转换比率输入部

103 元音转换部

104 子音声道信息保持部

105 子音选择部

106 子音变形部

107 合成部

111 被转换声音保持部

112 被转换声音信息发送部

113 目标元音声道信息发送部

114 被转换声音信息接收部

115 目标元音声道信息接收部

121 被转换声音服务器

122 目标声音服务器

201 目标讲话者声音

202 音素识别部

203 元音稳定区间抽取部

204 目标声道信息制作部

301 LPC 分析部

302 PARCOR 计算部

303 ARX 分析部

401 文本合成装置

具体实施方式

以下，参照附图来说明本发明的具体实施方式。

（实施例 1）

图 3 是本发明的实施例 1 所涉及的音质转换装置的框图。

实施例 1 所涉及的音质转换装置是通过根据被输入了输入声音的元音的声道信息的转换比率，来转换目标讲话者的元音的声道信息，从而转换输入声音的音质的装置，其包括：目标元音声道信息保持部 101、转换比率输入部 102、元音转换部 103、子音声道信息保持部 104、子音选择部 105、子音变形部 106、合成部 107。

目标元音声道信息保持部 101 是保持从目标讲话者发音的元音中抽取的声道信息的存储装置，例如，由硬盘或存储器等构成。

转换比率输入部 102 是输入进行音质转换时的向目标讲话者的转换比率的处理部。

元音转换部 103 是针对被输入了的附带音素边界信息的声道信息所包含的各个元音区间，根据由转换比率输入部 102 输入的转换比率，进行向附带音素边界信息的声道信息中的与目标元音声道信息保持部 101 所保持的该元音区间相对应的元音的声道信息的转换的处理部。另外，附带音素边界信息的声道信息是指，在输入声音的声道信息中附带了音素标记的信息。音素标记是指，包含与输入声音相对应的音素信息和各个音素的时间长度的信息的信息。关于附带音素边界信息的声道信息的生成方法，以后再述。

子音声道信息保持部 104 是保持，针对从多个讲话者的声音数据中抽取了的非特定讲话者的子音的声道信息的存储装置，例如，由硬盘或存储器等构成。

子音选择部 105 是根据附带音素边界信息的声道信息所包含的子音的声道信息的前后的元音的声道信息，从子音声道信息保持部 104 选择子音的声道信息的处理部，该子音的声道信息与通过元音转换部 103 元音的声道信息被变形后的附带音素边界信息的声道信息所包含的子音的声道信息相对应。

子音变形部 106 是，将由子音选择部 105 选择的子音的声道信息配合该子音的前后的元音的声道信息，进行变形的处理部。

合成部 107 是根据输入声音的声源信息和通过元音转换部 103、子音选择部 105 及子音变形部 106 被变形的附带音素边界信息的声道信息，合成声音的处理部。即，合成部 107 依据输入声音的声源信息生成激励声源，并驱动根据附带音素边界信息的声道信息而构成的声道滤波器从而合成声音。关于声源信息的生成方法，以后再述。

例如，音质转换装置由计算机等构成，并通过在计算机上执行程序来实现上述各个处理部。

其次，关于各自的构成部分进行详细地说明。

<目标元音声道信息保持部 101>

目标元音声道信息保持部 101 在是日语的情况下，保持目标讲话

者的至少五个元音(/aiueo/)的、来自目标讲话者的声道形状的声道信息。在是英语等其他语言的情况下，与日语的情况同样，关于各个元音保持声道信息即可。作为声道信息的表现方式，例如存在声道截面面积函数。声道截面面积函数表述如图 4(a)所示的，在以可变圆形截面面积的声管来模拟声道的声管模型中的各个声管的截面面积。众所周知，此截面面积与基于 LPC(Linear Predictive Coding：线性预测编码)分析的 PARCOR(Partial Auto Correlation：偏自相关)系数一一对应，并能够通过公式 1 来转换。在本实施例中，设通过 PARCOR 系数 ki 来表现声道信息。以后，虽然利用 PARCOR 系数来说明声道信息，但是，声道信息并不只限定于 PARCOR 系数，也可以利用与 PARCOR 系数等价的 LSP(Line Spectrum Pairs：线谱对)和 LPC 等。而且，所述声管模型中的声管之间的反射系数和 PARCOR 系数的关系，仅在于符号是相反的。因此，利用反射系数本身当然也没关系。

$$\frac{A_i}{A_{i+1}} = \frac{1-k_i}{1+k_i} \qquad \text{（公式 1）}$$

在此，A_n 表示如图 4(b)所示的第 i 区间的声管的截面面积，k_i 表示第 i 个和第 i+1 个边界的 PARCOR 系数(反射系数)。

利用根据 LPC 分析被分析出的线性预测系数 α_i，能够算出 PARCOR 系数。具体而言，通过利用 Levinson－Durbin－Itakura 算法，能够算出 PARCOR 系数。另外，PARCOR 系数具有以下的特征。

• 线性预测系数依赖于分析次数 p，而 PARCOR 系数则不依赖于分析的次数。

• 越是低次的系数，由于变动而对频谱的影响就越大，越成为高次则变动的影响就越小。

• 高次的系数的变动的影响平稳地涉及全部频带。

其次，关于目标讲话者的元音的声道信息（以下，称为“目标元

音声道信息”）的制作方法，边举例边进行说明。例如，目标元音声道信息能够通过由目标讲话者发出的孤立的元音声音来构筑。

图 5 是表示通过由目标讲话者发出的孤立的元音声音，生成目标元音声道信息保持部 101 所存储的目标元音声道信息的处理部的构成的图。

元音稳定区间抽取部 203 从被输入的孤立的元音声音中抽取孤立的元音的区间。抽取方法并不特别限定。例如，也可以将一定功率以上的区间作为稳定区间，并将该稳定区间作为元音的区间来抽取。

目标声道信息制作部 204 针对由元音稳定区间抽取部 203 抽取的元音的区间，算出上述 PARCOR 系数。

通过对发出被输入的孤立的元音的声音进行元音稳定区间抽取部 203 的处理以及目标声道信息制作部 204 的处理，从而构筑目标元音声道信息保持部 101。

在此之外，也可以通过如图 6 所示的处理部，构筑目标元音声道信息保持部 101。只要目标讲话者的发音至少包含五个元音，就并不限定于孤立的元音声音。例如，可以是目标讲话者临时自由发音后的声音，也可以是预先被收录的声音。另外，还可以利用歌唱数据等声音。

对这样的目标讲话者声音 201，音素识别部 202 进行音素识别。其次，元音稳定区间抽取部 203，根据在音素识别部 202 的识别结果，抽取稳定的元音区间。作为抽取的方法，例如，能够将在音素识别部 202 的识别结果的可靠性高的区间（似然高的区间）作为稳定的元音区间使用。

如此通过抽取稳定的元音区间，能够排除由于音素识别部 202 的识别错误的影响。例如，对关于在如图 7 所示的声音(/k//a//i/)被输入，并抽取元音区间/i/的稳定区间的情况进行说明。例如，能够将元音区间/i/内的功率大的区间设为稳定区间 50。或者，能够使用作为音素识别部 202 的内部信息的似然，将似然在阈值以上的区间作为稳定

区间来利用。

目标声道信息制作部 204 在抽出了的元音的稳定区间中，制作目标元音声道信息，并存储在目标元音声道信息保持部 101。通过此处理，能够构筑目标元音声道信息保持部 101。例如，由目标声道信息制作部 204 进行的目标元音声道信息的制作，通过计算前述的 PARCOR 系数来进行。

并且，目标元音声道信息保持部 101 所保持的目标元音声道信息的制作方法，并不限定于此，只要对稳定的元音区间进行声道信息的抽取，则也可以为其他的方法。

<转换比率输入部 102>

转换比率输入部 102 接受，对接近设为目标的讲话者的声音的程度进行指定的转换比率的输入。转换比率通常被指定为 0 以上 1 以下的数值。转换比率越接近 1，转换后的声音的音质就越接近目标讲话者，转换比率越接近 0，就越接近转换前声音的音质。

而且，通过输入 1 以上的转换比率，能够更加强调地表现转换前声音的音质和目标讲话者的音质之间的差别。并且，通过输入 0 以下的转换比率(负转换比率)，能够在相反的方向强调地表现转换前声音的音质和目标讲话者的音质之间的差别。另外，也可以省略转换比率的输入，将预先确定的比率作为转换比率来设定。

<元音转换部 103>

元音转换部 103 将被输入了的附带音素边界信息的声道信息所包含的元音区间的声道信息，以转换比率输入部 102 所指定的转换比率，转换为目标元音声道信息保持部 101 所保持的目标元音声道信息。以下对详细的转换方法进行说明。

通过从转换前的声音取得依据前述的 PARCOR 系数的声道信息，并通过将音素标记付与该声道信息，从而生成附带音素边界信息的声道信息。

具体如图 8A 所示，LPC 分析部 301 针对输入声音进行线性预测

分析，PARCOR 计算部 302 以分析后的线性预测系数为基础，算出 PARCOR 系数。并且，音素标记被另外付与。

而且，如下所述，求出输入合成部 107 的声源信息。即，逆滤波器部 304 从由 LPC 分析部 301 分析的滤波系数(线性预测系数)形成具备此频率响应的逆向特性的滤波器，并通过过滤输入声音，从而生成输入声音的声源波形(声源信息)。

也能够利用 ARX(autoregressive with exogenous input：外因输入自动回归)分析来代替上述 LPC 分析。ARX 分析是根据声音生成过程的声音分析法，该声音生成过程通过以高精度推定声道及声源参数为目的的 ARX 模式和声源模式公式来表述，该声音分析法与 LPC 分析相比，是能够高精度地分离声道信息和声源信息的声音分析法(非专利文献:大冢等「音源パルス列を考慮した頑健な ARX 音声分析法」(“考虑了声源脉冲串的强健的 ARX 声音分析法”)，日本声学学会会刊 58 卷 7 号（2002 年)、pp.386－397)。

图 8B 是表示附带音素边界信息的声道信息的其他的制作方法的图。

如该图所示，ARX 分析部 303 针对输入声音进行 ARX 分析，PARCOR 计算部 302 以分析后的全极点模型的多项式为基础，算出 PARCOR 系数。并且，音素标记被另外付与。

而且，输入合成部 107 的声源信息通过与图 8A 所示的在逆滤波器部 304 的处理相同的处理而生成。即，逆滤波器部 304 从由 ARX 分析部 303 分析的滤波系数形成具备此频率响应的逆向特性的滤波器，并通过过滤输入声音，从而生成输入声音的声源波形(声源信息)。

图 9 是表示附带音素边界信息的声道信息的另一其他的制作方法的图。

如图 9 所示，文本合成装置 401 从被输入的文本来合成声音，并输出合成声音。合成声音被输入 LPC 分析部 301 及逆滤波器部 304。因此，在输入声音是通过文本合成装置 401 合成的合成声音的情况下，

能够通过文本合成装置 401 取得音素标记。而且，LPC 分析部 301 及 PARCOR 计算部 302 通过利用合成后的声音，能够容易地算出 PARCOR 系数。

并且，输入合成部 107 的声源信息通过与图 8A 所示的在逆滤波器部 304 的处理相同的处理而生成。即，逆滤波器部 304 从由 ARX 分析部 303 分析的滤波系数形成具备此频率响应的逆向特性的滤波器，并通过过滤输入声音，从而生成输入声音的声源波形(声源信息)。

而且，在与音质转换装置脱机时生成附带音素边界信息的声道信息的情况下，也可以预先通过手动付与音素边界。

图 10A 至图 10J 是表示以十次 PARCOR 系数表现的元音/a/的声道信息的一个例子的图。

在该图中，纵座标轴表示反射系数，横坐标轴表示时间。从这些图中可以得知 PARCOR 系数针对时间变化进行比较平滑的变动。

元音转换部 103 如上所述，对被输入的附带音素边界信息的声道信息所包含的元音的声道信息进行转换。

首先，元音转换部 103 从目标元音声道信息保持部 101 取得与转换对象的元音的声道信息相对应的目标元音声道信息。在成为对象的目标元音声道信息为多个的情况下，元音转换部 103 配合成为转换对象的元音的音韵环境（例如前后的音素种类等）的状况，取得最合适的目标元音声道信息。

元音转换部 103 根据由转换比率输入部 102 输入的转换比率，将转换对象的元音的声道信息转换为目标元音声道信息。

在被输入的附带音素边界信息的声道信息中，根据公式 2 所示的多项式(第一函数)，将以成为转换对象的元音区间的 PARCOR 系数表现的声道信息的各因次的时间序列进行近似。例如，在十次 PARCOR 系数的情况下，各自的次数的 PARCOR 系数根据公式 2 所示的多项式来近似。因此，能够得出十种多项式。多项式的次数没有特别的限定，能够设定适当的次数。

$$\hat{y}_a = \sum_{i=0}^{p} a_i x^i \qquad \text{（公式 2）}$$

不过，$\hat{y}_a$ 是被输入的被转换声音的 PARCOR 系数的近似多项式，a_i 是多项式的系数，x 表示时刻。

此时作为适用多项式近似的单位，例如，能够将一个音素区间设为近似的单位。而且，也可以不是音素区间，而是设从音素中心到下一个音素中心为止的时间幅度为单位。另外，在以下的说明中，将音素区间作为单位来进行说明。

图 11A 至图 11D 是表示根据五次多项式对 PARCOR 系数进行近似，并以音素区间单位在时间方向上进行平滑化时的从一次至四次 PARCOR 系数的图。所谓图形的纵座标轴和横坐标轴，与图 10A 至图 10J 相同。

在本实施例中，作为多项式的次数虽然以五次为例进行了说明，但是多项式的次数也可以不是五次。并且，在根据多项式近似之外，也可以按每个音素区间根据回归线对 PARCOR 系数进行近似。

与成为转换对象的元音区间的 PARCOR 系数相同，根据公式 3 所示的多项式（第二函数），将以目标元音声道信息保持部 101 所保持的 PARCOR 系数来表现的目标元音声道信息进行近似，从而取得多项式的系数 bi。

$$\hat{y}_b = \sum_{i=0}^{p} b_i x^i \qquad \text{（公式 3）}$$

其次，利用被转换参数（a_i）、目标元音声道信息（b_i）、转换比率（r），根据公式 4 求出转换后的声道信息（PARCOR 系数）的多项式的系数 c_i 。

$$c_i = a_i + (b_i - a_i) \times r \qquad \text{（公式 4）}$$

通常，转换比率 r 在 0≤r≤1 的范围内被指定。但是，即使在转换比率 r 超出此范围的情况下，也能够根据公式 4 进行转换。在转换比率 r 超过 1 的情况下，成为更加强调被转换参数(a_i)与目标元音声道信息(b_i)之间的差分的转换。另一方面，在 r 是负值的情况下，成为在反方向上更加强调被转换参数(a_i)与目标元音声道信息(b_i)之间的差分的转换。

利用算出的转换后的多项式的系数 c_i ，以公式 5（第三函数）来求出转换后的声道信息。

$$\hat{y}_c = \sum_{i=0}^{p} c_i x^i \qquad \text{（公式 5）}$$

通过在 PARCOR 系数的各个因次中进行以上的转换处理，能够以被指定的转换比率转换为目标 PARCOR 系数。

图 12 表示实际上，针对元音/a/进行了上述转换的例子。在该图中，横坐标轴表示被归一化了的时间，纵座标轴表示第一次 PARCOR 系数。被归一化了的时间是指，通过以元音区间的持续时间长度，对时间进行归一化，从而具有从 0 到 1 为止的时刻的时间。这是在被转换声音的元音持续时间和目标元音声道信息的持续时间不相同的情况下，为了使时间轴一致的处理。图中的(a)表示被转换声音的男性讲话者的/a/的发音的系数的推移。同样，(b)表示目标元音的女性讲话者的/a/的发音的系数的推移。(c)表示利用上述转换方法，将男性讲话者的系数以转换比率 0.5 转换为女性讲话者的系数时的系数的推移。从该图可知，通过上述的变形方法，即能够对讲话者之间的 PARCOR 系数进行插值。

在音素边界为了防止 PARCOR 系数的值变得不连续，设置适当的过渡区间以进行插值处理。尽管插值的方法没有特别限定，但是例如能够通过进行线形插值来消除 PARCOR 系数的不连续。

图 13 是对关于设置过渡区间，对 PARCOR 系数的值进行插值的例子进行说明的图。该图表示元音/a/和元音/e/的连接边界的反射系数。在该图中的边界时刻(t)，反射系数变得不连续。于是，从边界时刻设置适当的过渡时间(Δt)，对从时刻 t-Δt 到时刻 t+Δt 之间的反射系数进行线形插值，通过求出插值后的反射系数 51 来防止在音素边界的反射系数的不连续。作为过渡时间，例如设为 20msec 即可。或者，也可以按照前后的元音继续时间长度来改变过渡时间。例如，也可以使元音区间越短过渡区间也就越短，元音区间越长过渡区间也就越长。

图 14A 是表示在对元音/a/和元音/i/的边界的 PARCOR 系数进行插值的情况下的频谱的图。图 14B 是表示将元音/a/和元音/i/的边界的声音通过平滑转换进行连接的情况下的频谱的图。在图 14A 及图 14B 中，纵座标轴表示频率，横坐标轴表示时间。在图 14A 中，可以得知，将在元音边界 21 的边界时刻作为 t 的情况下，在从时刻 t-Δt(22) 到时刻 t+Δt(23)为止的范围内，频谱上的强度的峰值为连续性变化。另一方面，在图 14B 中，频谱的峰值将元音边界 24 作为边界，不连续性地变化。如此通过对 PARCOR 系数的值进行插值，能够使频谱峰值(对应共振峰)连续性地变化。其结果为，由于共振峰连续性地变化，所以也能够使得到的合成音从/a/到/i/连续性地变化。

而且，图 15 是从将合成后的 PARCOR 系数进行了插值的 PARCOR 系数，再次抽取共振峰，并绘制的图。在该图中，纵座标轴表示频率(Hz)，横坐标轴表示时间(sec)。图上的点表示按每个合成音的帧的共振峰频率。附属在点上的竖棒表示共振峰的强度。竖棒越短共振峰强度就越强，竖棒越长共振峰强度就越弱。在以共振峰来看的情况下也可知，以元音边界 27 为中心的过渡区间（从时刻 28 到时刻 29 为止的区间）中，各个共振峰（共振峰强度也）连续性地变化。

如上所述，在元音边界中，通过设置适当的过渡区间，并对 PARCOR 系数进行插值，能够连续地转换共振峰及频谱，从而实现自然的音韵转变。

这样的频谱及共振峰的连续性的转变，在通过图 14B 所示的声音的平滑转换的连接中无法实现。

同样，图 16 是，在图 16(a)为/a/和/u/的连接，图 16(b)为/a/和/e/的连接，图 16(c)为/a/和/o/的连接之时的，表示根据平滑转换连接的频谱、对 PARCOR 系数进行插值后的频谱以及根据 PARCOR 系数插值的共振峰的移动的图。由此可知，在所有的元音连接中，能够使频谱强度的峰值连续地变化。

即表示了，通过进行以声道形状（PARCOR 系数）的插值，也能够进行共振峰的插值。因此，在合成音中也能够自然地表现元音的音韵转变。

图 17A 至图 17C 是表示在转换后的元音区间的时间上的中心的声道截面面积的图。此图是根据公式 1，将图 12 所示的在 PARCOR 系数的时间上的中心点的 PARCOR 系数转换为声道截面面积的图。在图 17A 至图 17C 的各个图形中，横坐标轴表示在声管中的位置，纵座标轴表示声道截面面积。图 17A 表示转换前的男性讲话者的声道截面面积，图 17B 表示目标讲话者的女性的声道截面面积，图 17C 表示以转换比率 50%，将转换前的 PARCOR 系数对应于转换后的 PARCOR 系数的声道截面面积。从这些图也可得知，图 17C 所示的声道截面面积为，转换前和转换后之间的中间的声道截面面积。

<子音声道信息保持部 104>

为了将音质转换为目标讲话者，虽然将在元音转换部 103 被输入的附带音素边界信息的声道信息所包含的元音转换为目标讲话者的元音声道信息，但是，由于转换元音，因而在子音和元音的连接边界上发生声道信息的不连续。

图 18 是在 VCV(V 表示元音，C 表示子音)音素列中，将元音转换部 103 进行元音的转换之后的某个 PARCOR 系数模式化表示的图。

在该图中，横坐标轴表示时间轴，纵座标轴表示 PARCOR 系数。图 18(a)是被输入的声音的声道信息。在此之中的元音部分的 PARCOR

系数利用图 18(b)所示的目标讲话者的声道信息，通过元音转换部 103 被变形。其结果为，得到如图 18(c)所示的元音部分的声道信息 10a 及 10b。但是，子音部分的声道信息 10c 未被转换，表示为输入声音的声道形状。因此，元音部分的声道信息和子音部分的声道信息之间的边界发生不连续性。因而关于子音部分的声道信息也需要转换。以下对关于子音部分的声道信息的转换方法进行说明。

声音的个人特性在考虑元音和子音的持续时间和稳定性等的情况下，可以考虑为主要根据元音来表现的。

于是，关于子音，能够不使用目标讲话者的声道信息，而从预先准备好的多个子音的声音信息之中，通过选择适合由元音转换部 103 转换后的元音声道信息的子音的声道信息，来缓和与转换后的元音在连接边界上的不连续性。在图 18(c)中，从子音声道信息保持部 104 所存储的子音的声道信息中，通过选择与前后的元音的声道信息 10a 及 10b 的连接性好的子音的声道信息 10d，能够实现缓和在音素边界上的不连续性。

为了实现以上的处理，预先从多个讲话者的多个发音中提出子音区间，与制作目标元音声道信息保持部 101 所存储的目标元音声道信息时同样，通过算出各个子音区间的 PARCOR 系数，来制作存储在子音声道信息保持部 104 的子音声道信息。

<子音选择部 105>

子音选择部 105 从子音声道信息保持部 104 选择，适合由元音转换部 103 转换了的元音声道信息的子音的声道信息。至于选择哪个子音声道信息，能够根据子音的种类(音素)和子音的始点及终点的连接点中的声道信息的连续性来判断。即，能够根据 PARCOR 系数的连接点中的连续性，来判断是否选择。具体而言，子音选择部 105 进行满足公式 6 的子音声道信息 C_i 的检索。

$$C_i = \arg\min_{C_k} [(w \times Cc(U_{i-1}, C_k) + (1-w) Cc(C_k, U_{i+1})] \quad \text{（公式 6）}$$

在此，U_{i-1}表示前面的音素的声道信息，U_{i+1}表示后续的音素的声道信息。

而且，w 是前面的音素与选择对象的子音之间的连续性和选择对象的子音与后续的音素之间的连续性的权重。权重 w 以重视与后续音素的连接的方式被适当地设定。之所以重视与后续音素的连接，是因为子音与后续的元音的结合比与前面的音素强。

并且，函数 Cc 是表示两个音素的声道信息的连续性的函数，例如，能够通过两个音素的边界上的 PARCOR 系数的差的绝对值来表现该连续性。而且，也可以设计成 PARCOR 系数越是低次的系数，权重就越大。

这样，通过选择适合向目标音质转换后的元音的声道信息的子音的声道信息，从而能够实现平滑的连接，并能够提高合成声音的自然性。

而且，还可以设计成仅设子音选择部 105 中选择的子音的声道信息为有声子音的声道信息，关于无声子音，使用被输入的声道信息。其理由是因为，无声子音是不伴随声带的振动的发音，声音的生成过程与生成元音或有声子音时不同。

<子音变形部 106>

虽然通过子音选择部 105，能够取得适合由元音转换部 103 转换后的元音声道信息的子音声道信息，但是，具有连接点的连续性并不一定充分的情况。因此，子音变形部 106 进行变形，以使由子音选择部 105 所选择了的子音的声道信息与后续元音的连接点能够连续地连接。

具体而言，子音变形部 106 使子音的 PARCOR 系数移动，以使在与后续元音的连接点中，PARCOR 系数和后续元音的 PARCOR 系数一致。但是，为了保证稳定性，PARCOR 系数必须在[－1,1]的范围内。因此，暂且根据 tanh－1 函数等将 PARCOR 系数映射在[－∞,

∞]的空间中，并在映射后的空间上进行线性移动之后，通过再次根据tanh返回[－1,1]的范围，从而能够既保证了稳定性，又改善子音区间与后续元音区间的声道形状的连续性。

<合成部107>

合成部107利用音质转换后的声道信息和另外被输入的声源信息来合成声音。虽然没有特别限定合成的方法，但是，在利用PARCOR系数作为声道信息的情况下，利用PARCOR合成即可。或者，也可以在从PARCOR系数转换成LPC系数之后合成声音，还可以从PARCOR系数中抽取共振峰，通过共振峰合成来合成声音。进而，也可以从PARCOR系数算出LSP系数，通过LSP合成来合成声音。

其次，关于在本实施例中被执行的处理，利用图19A及图19B所示的流程图进行说明。

在本发明的实施例中被执行的处理大致由两个处理组成。一个是目标元音声道信息保持部101的构筑处理，另一个是音质的转换处理。

首先，参照图19A对有关目标元音声道信息保持部101的构筑处理进行说明。

从目标讲话者发出了的声音中抽取元音的稳定区间(步骤S001)。作为稳定区间的抽取方法，如上所述，音素识别部202识别音素，元音稳定区间抽取部203将在识别结果中所包含的元音区间之中的似然是阈值以上的元音区间作为元音稳定区间来抽取。

目标声道信息制作部204制作在被抽取的元音区间中的声道信息(步骤S002)。如上所述，声道信息能够通过PARCOR系数来表示。PARCOR系数能够从全极点模型的多项式中算出。因此，作为分析方法，能够使用LPC分析或者ARX分析。

目标声道信息制作部204将在步骤S002中被分析了的元音稳定区间的PARCOR系数作为声道信息，登记在目标元音声道信息保持部101(步骤S003)。

通过以上步骤，能够构筑对针对目标讲话者的音质附加特征的目

标元音声道信息保持部 101。

其次，参照图 19B 对有关通过图 3 所示的音质转换装置，将被输入的附带音素边界信息的声音转换为目标讲话者的声音的处理进行说明。

转换比率输入部 102 接受表示向目标讲话者的转换的程度的转换比率的输入(步骤 S004)。

元音转换部 103 针对被输入的声音的元音区间，从目标元音声道信息保持部 101 取得针对所对应的元音的目标声道信息，根据在步骤 S004 中被输入的转换比率，转换被输入的声音的元音区间的声道信息(步骤 S005)。

子音选择部 105 选择适合被转换了的元音区间的声道信息的子音声道信息(步骤 S006)。此时，设子音选择部 105 将子音的种类（音素）以及子音与其前后的音素的连接点中的声道信息的连续性作为评价的标准，选择连续性最高的子音的声道信息。

为了提高被选择的子音的声道信息与在前后的音素区间的声道信息的连续性，子音变形部 106 将子音的声道信息进行变形(步骤 S007)。根据被选择的子音的声道信息与前后的音素区间的各自的连接点中的声道信息(PARCOR 系数)的差分值，通过使子音的 PARCOR 系数移动来实现变形。并且，在使之移动之时，为了保证 PARCOR 系数的稳定性，根据 tanh－1 函数等，暂且将 PARCOR 系数映射在[－∞,∞]的空间，并在映射后的空间中线性移动 PARCOR 系数，移动后再次根据 tanh 函数等返回[－1,1]的空间。因此，能够进行稳定后的子音声道信息的转换。并且，从[－1,1]向[－∞,∞]的映射不仅限于 tanh－1 函数，也可以利用 f(x)=sgn(x)×1/(1－|x|)等的函数。在此，sgn(x)是在 x 为正的时候成为+1，在 x 为负的时候成为－1 的函数。

这样，通过对子音区间的声道信息进行变形，能够制作适合转换后的元音区间且连续性高的子音区间的声道信息。因此，能够实现稳定连续的，且为高音质的音质转换。

合成部 107 根据通过元音转换部 103、子音选择部 105 以及子音变形部 106 所转换了的声道信息，生成合成音(步骤 S008)。此时，作为声源信息，能够使用转换前声音的声源信息。通常，在 LPC 系统的分析合成中，因为作为激励声源使用脉冲串的情况较多，所以也可以根据预先设定了的基频等信息，在对声源信息(F0(基频)、功率等)进行变形之后，生成合成音。因此，不仅能够进行依据声道信息的声调的转换，也能够进行依据基频等表示的韵律、或者声源信息的转换。

而且，例如在合成部 107 中，也能够使用 Rosenberg－Klatt 模型等的声门声源模型，在使用了这样的构成的情况下，还能够使用利用从被转换声音的 Rosenberg－Klatt 模型的参数(OQ、TL、AV、F0 等)向目标声音移动后的值等方法。

根据所涉及的构成，将附带音素边界信息的声道信息作为输入，元音转换部 103 根据由转换比率输入部 102 输入的转换比率，进行从被输入了的附带音素边界信息的声道信息所包含的各个元音区间的声道信息，向与目标元音声道信息保持部 101 所保持的该元音区间相对应的元音的声道信息的转换。子音选择部 105 根据子音的前后的元音的声道信息，从子音声道信息保持部 104 选择适合由元音转换部 103 转换了的元音声道信息的子音的声道信息。子音变形部 106 将由子音选择部 105 选择的子音的声道信息配合前后的元音的声道信息，来进行变形。合成部 107 根据通过元音转换部 103、子音选择部 105 以及子音变形部 106 变形了的附带音素边界信息的声道信息，合成声音。因此，作为目标讲话者的声道信息，只准备元音稳定区间的声道信息即可。而且，在制作目标讲话者的声道信息之时，由于只需识别元音稳定区间即可，所以不受如专利文献 2 的技术那样的声音识别错误的影响。

即，由于能够大大减少针对目标讲话者的负担，所以能够容易地进行音质转换。而且，在专利文献 2 的技术中，根据在声音合成部 14 的声音合成中所使用的声音单元与目标讲话者的发音之间的差分，来

编写变换函数。因此，被转换声音的音质必须与声音合成用数据存储部 13 所保持的声音单元的音质相同或者类似。对此，本发明的音质转换装置将目标讲话者的元音声道信息作为绝对的目标。为此，可以完全不限制转换前的声音的音质，从而可以输入任何音质的声音。即，因为对被输入的被转换声音的限制非常少，所以能够针对广泛的声音进行该声音的音质的转换。

同时，通过子音选择部 105 从子音声道信息保持部 104 选择预先被保持的子音的声道信息，从而能够使用适合转换后的元音的声道信息的最佳的子音声道信息。

而且，在本实施例中，通过子音选择部 105 及子音变形部 106，不仅在元音区间，并且在子音区间中也进行了转换声源信息的处理，但是，也可以省略这些处理。在此情况下，作为子音的声道信息，就照原样使用被输入音质转换装置的附带音素边界信息的声道信息所包含的子音的声道信息。因此，无论在处理终端的处理性能低的情况下，或在存储容量少的情况下，都能够实现向目标讲话者的音质转换。

另外，也可以只省略子音变形部 106，而构成音质转换装置。在此情况下，就照原样使用子音选择部 105 所选择的子音的声道信息。

或者，也可以只省略子音选择部 105，而构成音质转换装置。在此情况下，子音变形部 106 对被输入音质转换装置的附带音素边界信息的声道信息所包含的子音的声道信息进行变形。

（实施例 2）

以下，对本发明的实施例 2 进行说明。

实施例 2 与实施例 1 的音质转换装置不同，考虑的是被转换声音和目标音质信息被个别地管理的情况。并考虑被转换声音是声音内容。例如，是唱歌声音等。作为目标音质信息，设为保持着各种各样的音质。例如，设为保持着各种各样的歌手的音质信息。在这种情况下，可以考虑分别下载声音内容和目标音质信息，从而在终端进行音质转换的使用方法。

图 20 是表示本发明的实施例 2 所涉及的音质转换系统的构成的图。关于图 20 中的与图 3 相同的构成部分使用同样的符号，并省略对其的说明。

音质转换系统包括：被转换声音服务器 121、目标声音服务器 122、终端 123。

被转换声音服务器 121 是管理并提供被转换声音信息的服务器，包括：被转换声音保持部 111、被转换声音信息发送部 112。

被转换声音保持部 111 是保持被转换声音的信息的存储装置，例如，由硬盘或存储器等构成。

被转换声音信息发送部 112 是将被转换声音保持部 111 所保持的被转换声音信息，通过网络发送到终端 123 的处理部。

目标声音服务器 122 是管理并提供成为目标的音质信息的服务器，包括：目标元音声道信息保持部 101、目标元音声道信息发送部 113。

目标元音声道信息发送部 113 是将目标元音声道信息保持部 101 所保持的目标讲话者的元音声道信息，通过网络发送到终端 123 的处理部。

终端 123 是根据从目标声音服务器 122 发送的目标元音声道信息，对从被转换声音服务器 121 发送的被转换声音信息的音质进行转换的终端装置，包括：被转换声音信息接收部 114、目标元音声道信息接收部 115、转换比率输入部 102、元音转换部 103、子音声道信息保持部 104、子音选择部 105、子音变形部 106、合成部 107。

被转换声音信息接收部 114 是通过网络，接收由被转换声音信息发送部 112 发送了的被转换声音信息的处理部。

目标元音声道信息接收部 115 是通过网络，接收由目标元音声道信息发送部 113 发送了的目标元音声道信息的处理部。

被转换声音服务器 121、目标声音服务器 122 以及终端 123，例如，由具备 CPU、存储器、通信接口等的计算机等构成，上述各个处

理部通过在计算机的 CPU 上执行程序来实现。

本实施例与实施例 1 的不同之处在于，作为目标讲话者的元音的声道信息的目标元音声道信息和作为与被转换声音对应的信息的被转换声音信息，通过网络进行收发。

其次，关于实施例 2 所涉及的音质转换系统的工作进行说明。图 21 是表示本发明的实施例 2 所涉及的音质转换系统的处理的流程的流程图。

终端 123 通过网络，对目标声音服务器 122 请求目标讲话者的元音声道信息。目标声音服务器 122 的目标元音声道信息发送部 113 从目标元音声道信息保持部 101 取得被请求了的目标讲话者的元音声道信息，并发送到终端 123。终端 123 的目标元音声道信息接收部 115 接收目标讲话者的元音声道信息(步骤 S101)。

并不特别限定目标讲话者的指定方法，例如也可以利用讲话者标识符进行指定。

终端 123 通过网络，对被转换声音服务器 121 请求被转换声音信息。被转换声音服务器 121 的被转换声音信息发送部 112 从被转换声音保持部 111 取得被请求了的被转换声音信息，并发送到终端 123。终端 123 的被转换声音信息接收部 114 接收被转换声音信息(步骤 S102)。

并不特别限定被转换声音信息的指定方法，例如也可以通过标识符来管理声音内容，并利用此标识符进行指定。

转换比率输入部 102 接受表示向目标讲话者的转换的程度的转换比率的输入(步骤 S004)。另外，也可以省略转换比率的输入，而设定预先确定的转换比率。

元音转换部 103 针对被输入的声音的元音区间，从目标元音声道信息接收部 115 取得对应的元音的目标元音声道信息，根据在步骤 S004 中被输入的转换比率，转换被输入的声音的元音区间的声道信息(步骤 S005)。

子音选择部 105 选择适合被转换了的元音区间的声道信息的子音声道信息(步骤 S006)。此时，设子音选择部 105 将子音与其前后的音素的连接点中的声道信息的连续性作为评价标准，并选择连续性最高的子音的声道信息。

为了提高被选择的子音的声道信息与在前后的音素区间的声道信息的连续性，子音变形部 106 将子音的声道信息进行变形(步骤 S007)。根据被选择的子音的声道信息与前后的音素区间的各自的连接点中的声道信息(PARCOR 系数)的差分值，通过使子音的 PARCOR 系数移动来实现变形。并且，在使之移动之时，为了保证 PARCOR 系数的稳定性，根据 tanh－1 函数等，暂且将 PARCOR 系数映射在[－∞,∞]的空间，并在映射后的空间中线性移动 PARCOR 系数，移动后再次根据 tanh 函数等返回[－1,1]的空间。因此，能够进行稳定后的子音声道信息的转换。并且，从[－1,1]向[－∞,∞]的映射不仅限于 tanh－1 函数，也可以利用 f(x)=sgn(x)×1/(1－|x|)等的函数。在此，sgn(x)是在 x 为正的时候成为+1，在 x 为负的时候成为－1 的函数。

这样，通过对子音区间的声道信息进行变形，能够制作适合转换后的元音区间且连续性高的子音区间的声道信息。因此，能够实现稳定连续的，且为高音质的音质转换。

合成部 107 根据通过元音转换部 103、子音选择部 105 以及子音变形部 106 所转换了的声道信息，生成合成音(步骤 S008)。此时，作为声源信息，能够使用转换前声音的声源信息。并且，也可以在根据预先设定了的基频等信息将声源信息进行变形之后，生成合成音。因此，不仅能够进行依据声道信息的声调的转换，也能够进行依据基频等表示的韵律、或者声源信息的转换。

另外，也可以不是步骤 S101、步骤 S102、步骤 S004 的顺序，可以按任意的顺序来执行。

根据所涉及的构成，目标声音服务器 122 管理并发送目标声音信息。因此，不需要在终端 123 制作目标声音信息，且能够进行向登记

在目标声音服务器 122 上的各种各样的音质的音质转换。

并且，通过被转换声音服务器 121 管理并发送被转换声音，从而不需要在终端 123 制作被转换声音信息，就能够利用登记在被转换声音服务器 121 上的各种各样的被转换声音信息。

通过被转换声音服务器 121 管理声音内容，目标声音服务器 122 管理目标讲话者的音质信息，从而能够分别管理声音信息和讲话者的音质信息。因此，终端 123 的使用者能够以适合自己的爱好的音质来收听适合自己的爱好的声音内容。

例如，通过以被转换声音服务器 121 管理唱歌声音，并以目标声音服务器 122 管理各种各样的歌手的目标声音信息，能够在终端 123 中将各种各样的音乐转换成各种各样的歌手的音质来收听，从而能够提供适合使用者的爱好的音乐。

并且，也可以通过同一个服务器来实现被转换声音服务器 121 和目标声音服务器 122。

（实施例 3）

在实施例 2 表示了使用服务器来管理被转换声音和目标元音声道信息，且终端分别将其下载，以生成转换了音质的声音的利用方法。对此，在本实施例中，用户利用终端来登记自己的声音的音质，例如，对将本发明适用于，将用于对用户通知来电呼叫的来电歌声等转换为自己的音质来享受的服务的情况进行说明。

图 22 是表示本发明的实施例 3 所涉及的音质转换系统的构成的图。关于图 22 中的与图 3 相同的构成部分使用同样的符号，并省略对其的说明。

音质转换系统包括：被转换声音服务器 121、音质转换服务器 222、终端 223。

被转换声音服务器 121 具有与实施例 2 所示的被转换声音服务器 121 相同的构成，包括：被转换声音保持部 111、被转换声音信息发送部 112。但是，依据被转换声音信息发送部 112 的被转换声音信息

的发送目的地不同，本实施例所涉及的被转换声音信息发送部 112 通过网络，将被转换声音信息发送到音质转换服务器 222。

终端 223 是为了用户享受歌声转换服务的终端装置。即，终端 223 是制作成为目标的音质信息，并将其提供给音质转换服务器 222，且接收并再生由音质转换服务器 222 转换了的歌声声音的装置，包括：声音输入部 109、目标元音声道信息制作部 224、目标元音声道信息发送部 113、被转换声音指定部 1301、转换比率输入部 102、音质转换声音接收部 1304、再生部 305。

声音输入部 109 是为了取得用户的声音的装置，例如，包括扩音器等。

目标元音声道信息制作部 224 是制作，作为从目标讲话者、即声音输入部 109 输入了声音的用户的元音的声道信息的目标元音声道信息的处理部。并不限定目标元音声道信息的制作方法，例如，目标元音声道信息制作部 224 根据图 5 所示的方法制作目标元音声道信息，并包括元音稳定区间抽取部 203、目标声道信息制作部 204。

目标元音声道信息发送部 113 是通过网络，将由目标元音声道信息制作部 224 制作了的目标元音声道信息发送到音质转换服务器 222 的处理部。

被转换声音指定部 1301 是，从被转换声音服务器 121 所保持的被转换声音信息中，指定作为转换对象的被转换声音信息，并将指定的结果通过网络发送到音质转换服务器 222 的处理部。

虽然转换比率输入部 102 具有与实施例 1 及 2 所示的转换比率输入部 102 同样的构成，但是，本实施例所涉及的转换比率输入部 102 还通过网络，将被输入了的转换比率发送到音质转换服务器 222。另外，也可以省略转换比率的输入，而使用预先确定的转换比率。

音质转换声音接收部 1304 是接收作为通过音质转换服务器 222，音质被转换了的被转换声音的合成音的处理部。

再生部 306 是再生音质转换声音接收部 1304 所接收了的合成音

的装置，例如，包括扬声器等。

音质转换服务器222是根据从终端223的目标元音声道信息发送部113发送的目标元音声道信息，对从被转换声音服务器121发送的被转换声音信息的音质进行转换的终端装置，包括：被转换声音信息接收部114、目标元音声道信息接收部115、转换比率接收部1302、元音转换部103、子音声道信息保持部104、子音选择部105、子音变形部106、合成部107、合成声音发送部1303。

转换比率接收部1302是接收从转换比率输入部102发送了的转换比率的处理部。

合成声音发送部1303是通过网络，将由合成部107输出的合成音发送到终端223的音质转换声音接收部1304的处理部。

被转换声音服务器121、音质转换服务器222以及终端223，例如，由具备CPU、存储器、通信接口等的计算机等构成，上述各个处理部通过在计算机的CPU上执行程序来实现。

本实施例与实施例2的不同之处在于，通过终端223在抽取了成为目标的音质特征之后，将其发送到音质转换服务器222，并通过音质转换服务器222将音质转换后的合成音送回到终端223，从而能够在终端223上得到具有抽取了的音质特征的合成音。

其次，关于实施例3所涉及的音质转换系统的工作进行说明。图23是表示本发明的实施例3所涉及的音质转换系统的处理的流程的流程图。

终端223利用声音输入部109，取得用户的元音声音。例如，能够通过用户对着扩音器进行“あ、い、う、え、お”的发音，来取得元音声音。元音声音的取得方法并不仅限于此，还可以如图6所示，从被发音了的文章中抽取元音声音(步骤S301)。

终端223根据利用目标元音声道信息制作部224取得了的元音声音，制作声道信息。声道信息的制作方法可以与实施例1相同(步骤S302)。

终端 223 利用被转换声音指定部 1301，指定被转换声音信息。指定的方法并不特别限定。被转换声音服务器 121 的被转换声音信息发送部 112 从被转换声音保持部 111 所保持的被转换声音信息之中，选择由被转换声音指定部 1301 指定了的被转换声音信息，并将所选择的被转换声音信息发送到音质转换服务器 222(步骤 S303)。

终端 223 利用转换比率输入部 102 来取得转换比率(步骤 S304)。

音质转换服务器 222 的转换比率接收部 1302 接收由终端 223 发送的转换比率，目标元音声道信息接收部 115 接收由终端 223 发送的目标元音声道信息。而且，被转换声音信息接收部 114 接收由被转换声音服务器 121 发送的被转换声音信息。并且，元音转换部 103 针对接收了的被转换声音信息的元音区间的声道信息，从目标元音声道信息接收部 115 取得对应的元音的目标元音声道信息，根据由转换比率接收部 1302 接收了的转换比率，转换元音区间的声道信息(步骤 S305)。

音质转换服务器 222 的子音选择部 105 选择适合被转换了的元音区间的声道信息的子音声道信息(步骤 S306)。此时，设子音选择部 105 将子音与其前后的音素的连接点中的声道信息的连续性作为评价的标准，并选择连续性最高的子音的声道信息。

音质转换服务器 222 的子音变形部 106 为了提高被选择的子音的声道信息与在前后的音素区间的连续性，对子音的声道信息进行变形(步骤 S307)。

作为变形的方法，可以与实施例 2 的变形方法相同。这样，通过对子音区间的声道信息进行变形，能够制作适合转换后的元音区间且连续性高的子音区间的声道信息。因此，能够实现稳定连续的，且为高音质的音质转换。

音质转换服务器 222 的合成部 107 根据通过元音转换部 103、子音选择部 105 以及子音变形部 106 转换了的声道信息，生成合成音，合成声音发送部 1303 将生成的合成音发送到终端 223(步骤 S308)。

此时，作为合成音生成时的声源信息，能够使用转换前声音的声源信息。并且，也可以在根据预先设定了的基频等信息将声源信息进行变形之后，生成合成音。因此，不仅能够进行依据声道信息的声调的转换，也能够进行依据基频等表示的韵律、或者声源信息的转换。

终端 223 的音质转换声音接收部 1304 接收由合成声音发送部 1303 发送的合成音，再生部 305 再生接收了的合成音(S309)。

根据所涉及的构成，终端 223 制作并发送目标声音信息，且接收并再生通过音质转换服务器 222 转换了音质的声音。因此，只需在终端 223 输入成为目标的声音，并制作成为目标的元音的声道信息即可，从而能够大为减小终端 223 的处理负荷。

而且，通过利用被转换声音服务器 121 来管理被转换声音信息，并通过将被转换声音信息从被转换声音服务器 121 发送到音质转换服务器 222，从而不需要在终端 223 制作被转换声音信息。

因为被转换声音服务器 121 管理声音内容，且终端 223 只制作成为目标的音质，所以终端 223 的使用者能够以适合自己的爱好的音质来收听适合自己的爱好的声音内容。

例如，通过以被转换声音服务器 121 管理唱歌声音，并通过利用音质转换服务器 222 将唱歌声音转换成由终端 223 取得了的目标音质，从而能够提供适合使用者的爱好的音乐。

并且，也可以通过同一个服务器来实现被转换声音服务器 121 和音质转换服务器 222。

作为本实施例的应用例，例如在终端 223 是移动电话的情况下，例如通过将取得了的合成音作为铃声来登记，从而用户能够制作自己独有的铃声。

并且，在本实施例的构成中，由于以音质转换服务器 222 来进行音质转换，所以能够以服务器来进行音质转换的管理。因此，能够管理用户的音质转换的履历，具有不容易发生侵犯版权及肖像权的问题之效果。

另外，在本实施例中，虽然目标元音声道信息制作部 224 被设置在终端 223，但是，也可以设置在音质转换服务器 222。在此情况下，通过网络，将声音输入部 109 所输入的目标元音声音发送到音质转换服务器 222。而且，在音质转换服务器 222 中，还可以利用目标元音声道信息制作部 224 从接收了的声音制作目标元音声道信息，并在通过元音转换器 103 的音质转换之时使用。根据此构成，由于终端 223 只需输入成为目标的音质的元音即可，因此具有大为减小处理负荷的效果。

再者，本实施例不仅能够适用于移动电话的来电歌声的音质转换，例如，通过以用户的音质来再生歌手所演唱的歌曲，从而能够收听既具备专业歌手的演唱能力又是以用户的音质来演唱的歌曲。由于通过模仿歌曲的演唱从而能够学习专业歌手的演唱能力，因此能够适用于卡拉 OK 的练习等用途上。

在此公开的实施例的所有部分都是例示，应当认为并不是加以限制的内容。本发明的范围不在于上述的说明，是根据权利要求而表示的，并意味着包括与权利要求同等的意思以及在范围内的所有变更。

本发明所涉及的音质转换装置具有从目标讲话者的元音区间的声道信息，高品质地转换音质的功能，作为需要各种各样的音质的用户界面或娱乐等非常有用。并且，能够应用于通过移动电话等的声音通信中的语音变换器等用途上。

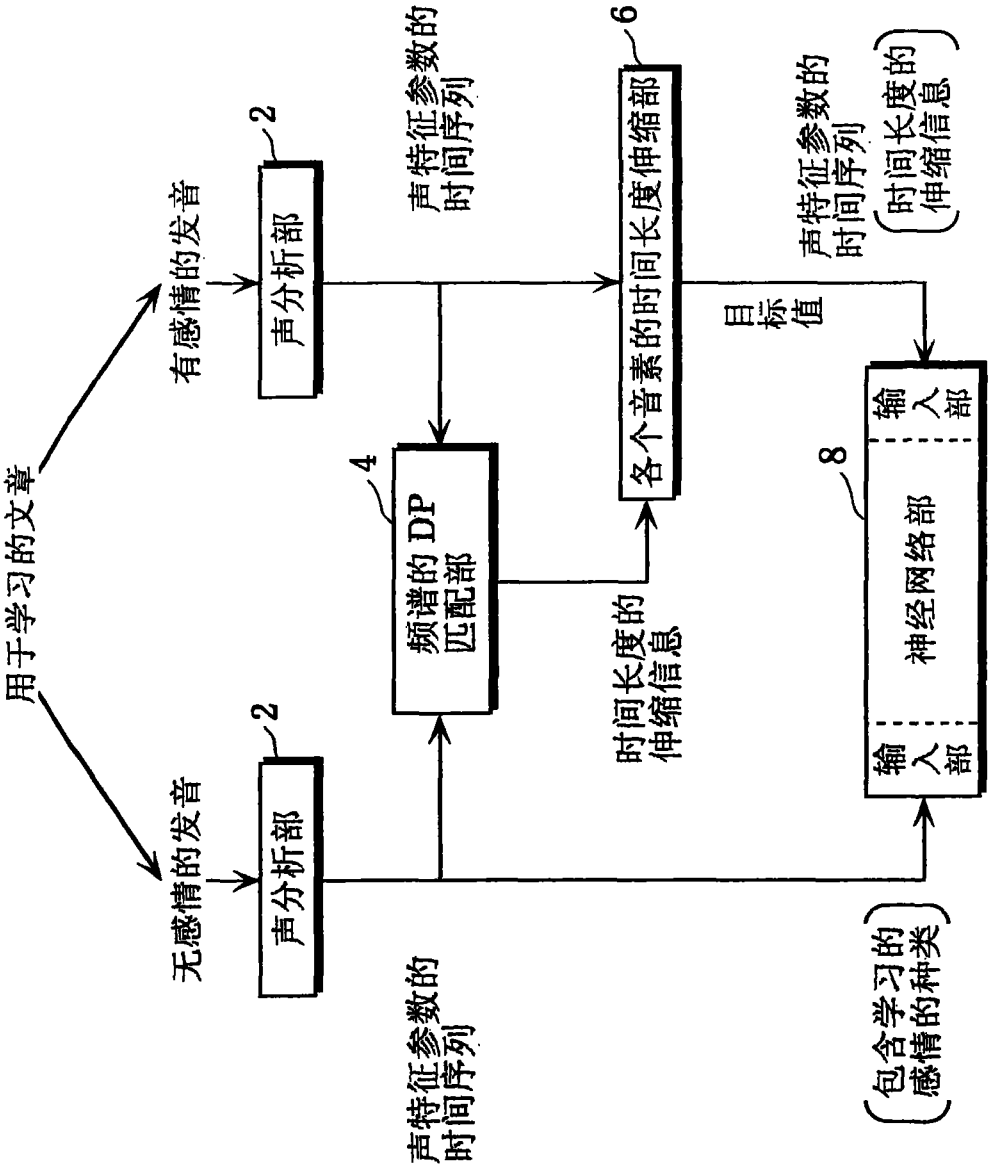

用于学习的文章
有感情的发音
无感情的发音
声分析部
2
声分析部
2
声特征参数的时间序列
声特征参数的时间序列
频谱的 DP 匹配部
4
各个音素的时间长度伸缩部
6
时间长度的伸缩信息
目标值
声特征参数的时间序列
时间长度的伸缩信息
神经网络部
8
输入部
输入部
包含学习的感情的种类

图 1

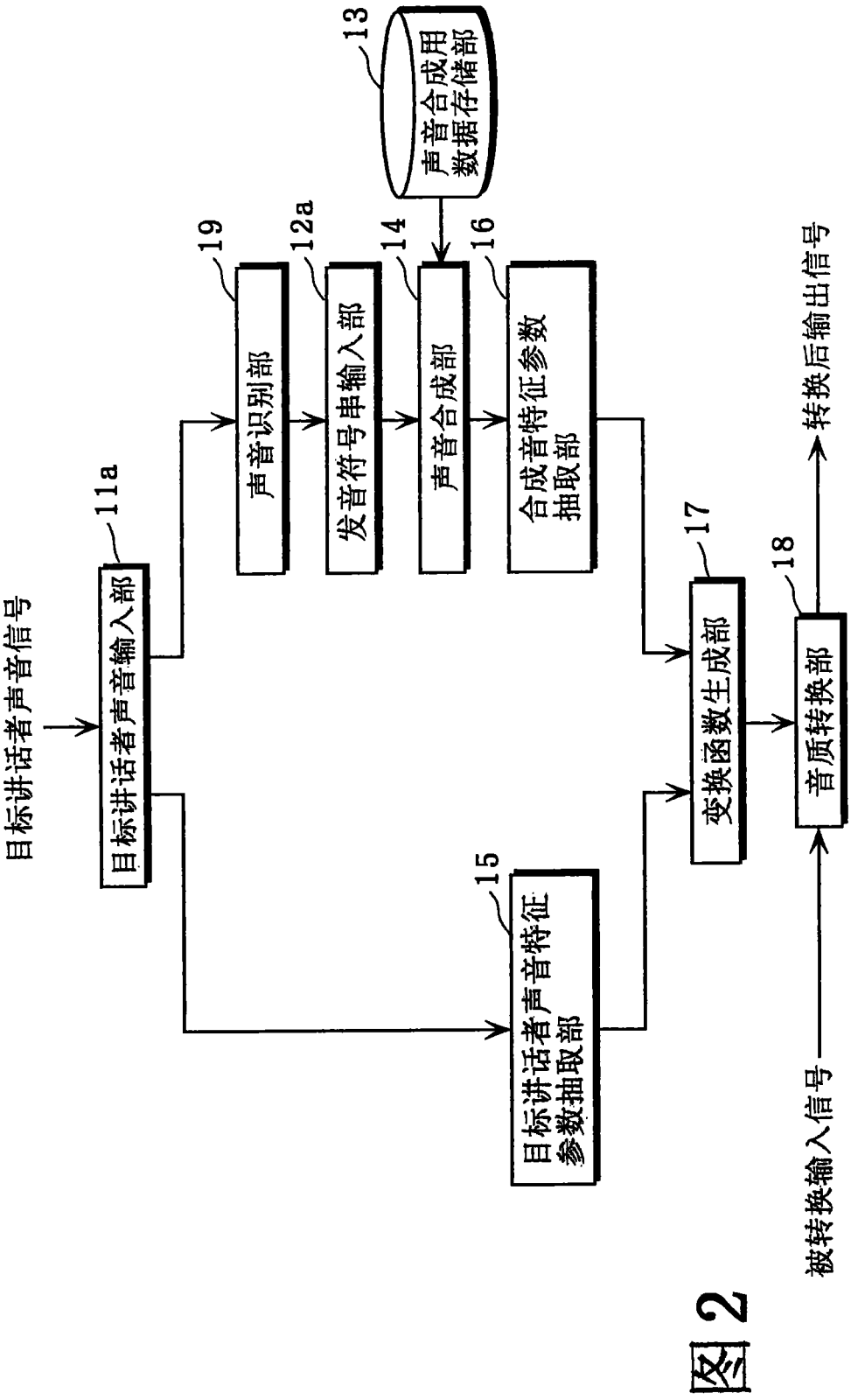

目标讲话者声音信号
目标讲话者声音输入部
11a
声音识别部
19
发音符号串输入部
12a
声音合成用数据存储部
13
声音合成部
14
合成音特征参数抽取部
16
目标讲话者声音特征参数抽取部
15
变换函数生成部
17
音质转换部
18
转换后输出信号
被转换输入信号

图 2

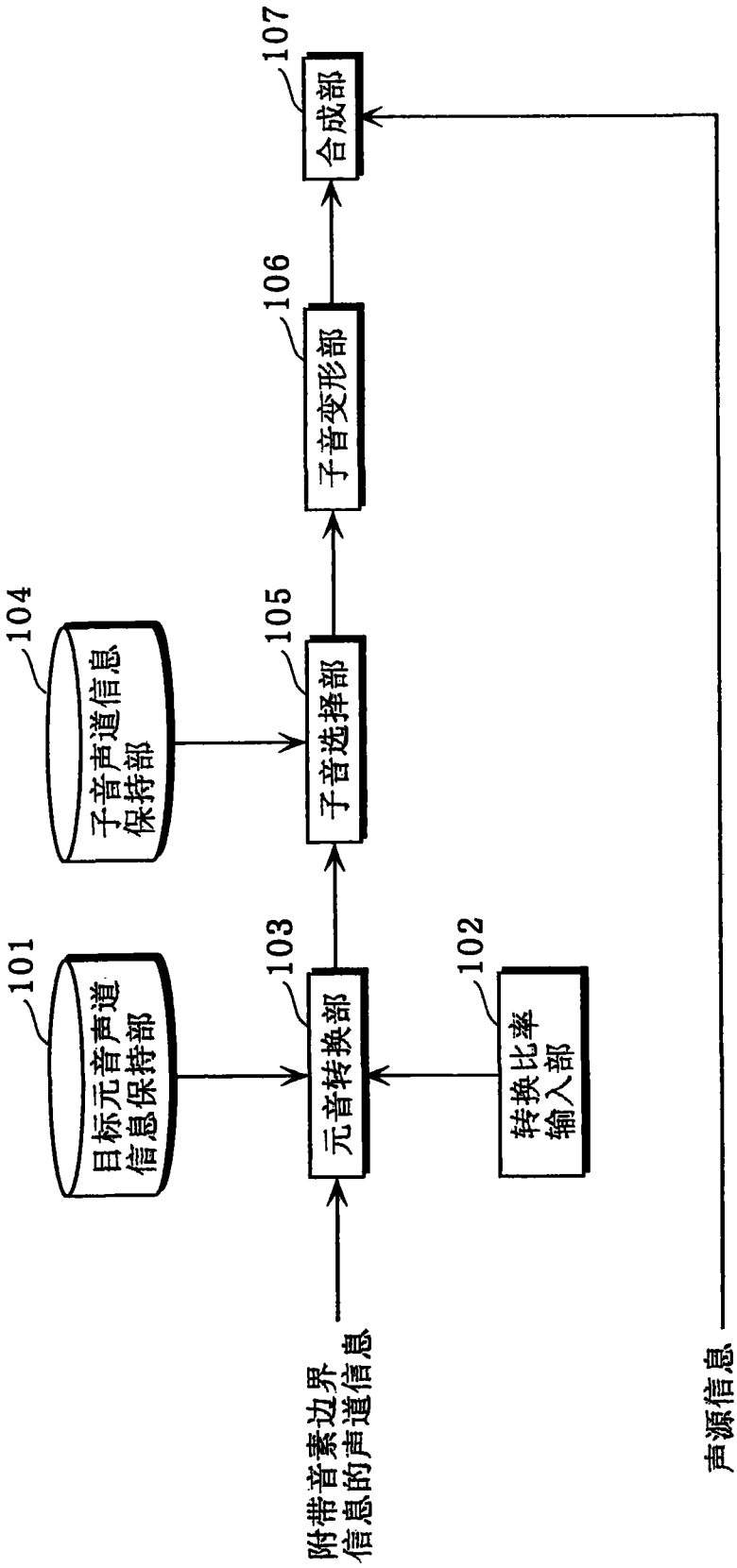

附带音素边界信息的声道信息
声源信息
元音转换部
转换比率输入部
子音选择部
子音变形部
合成部
目标元音声道信息保持部
子音声道信息保持部
101
102
103
104
105
106
107

图3

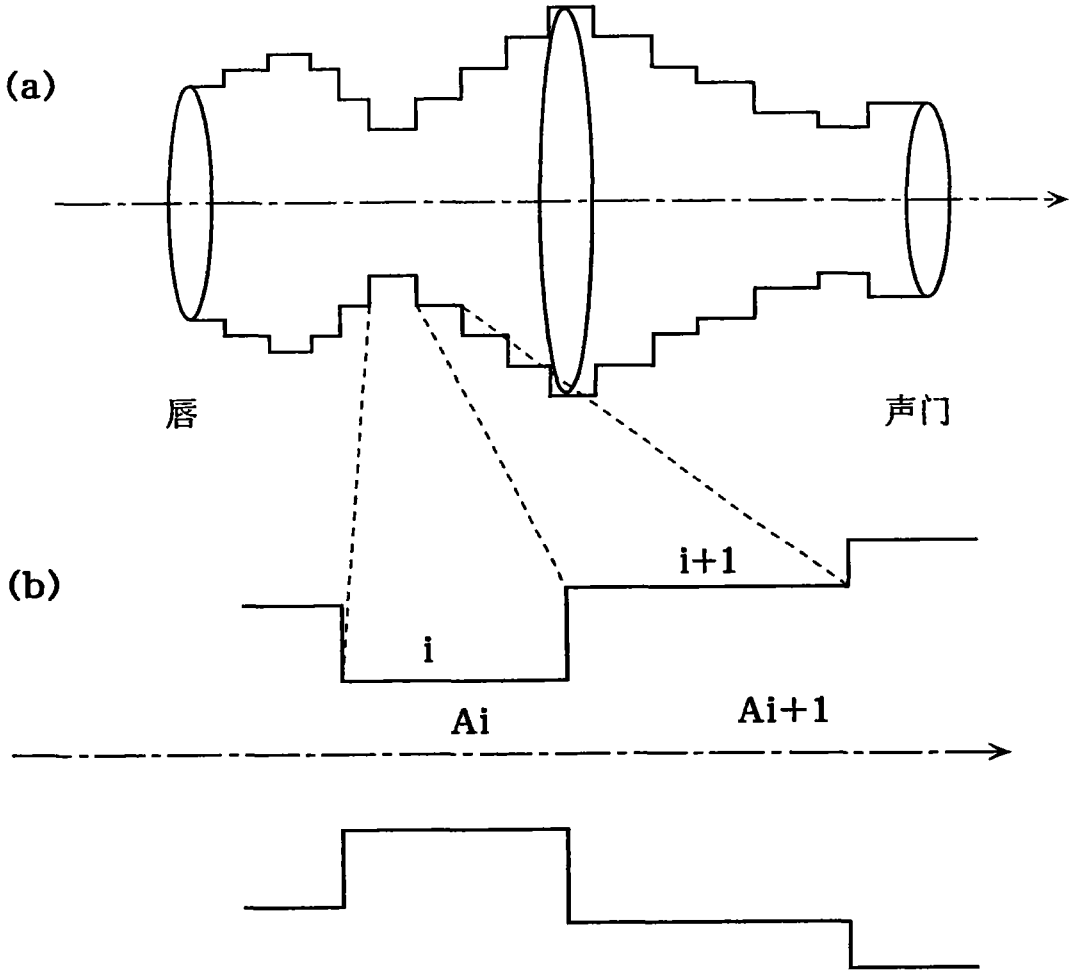


图 4

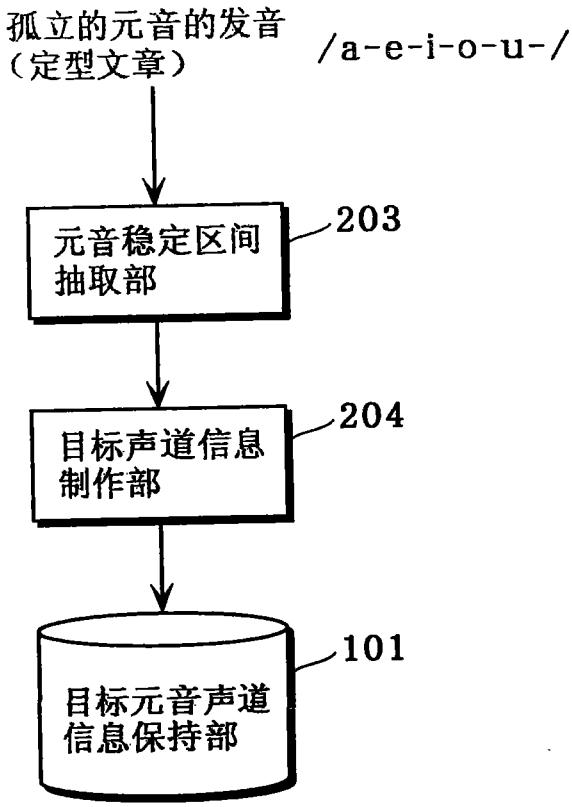

孤立的元音的发音
（定型文章）
/a-e-i-o-u-/
元音稳定区间
抽取部
203
目标声道信息
制作部
204
目标元音声道
信息保持部
101

图5

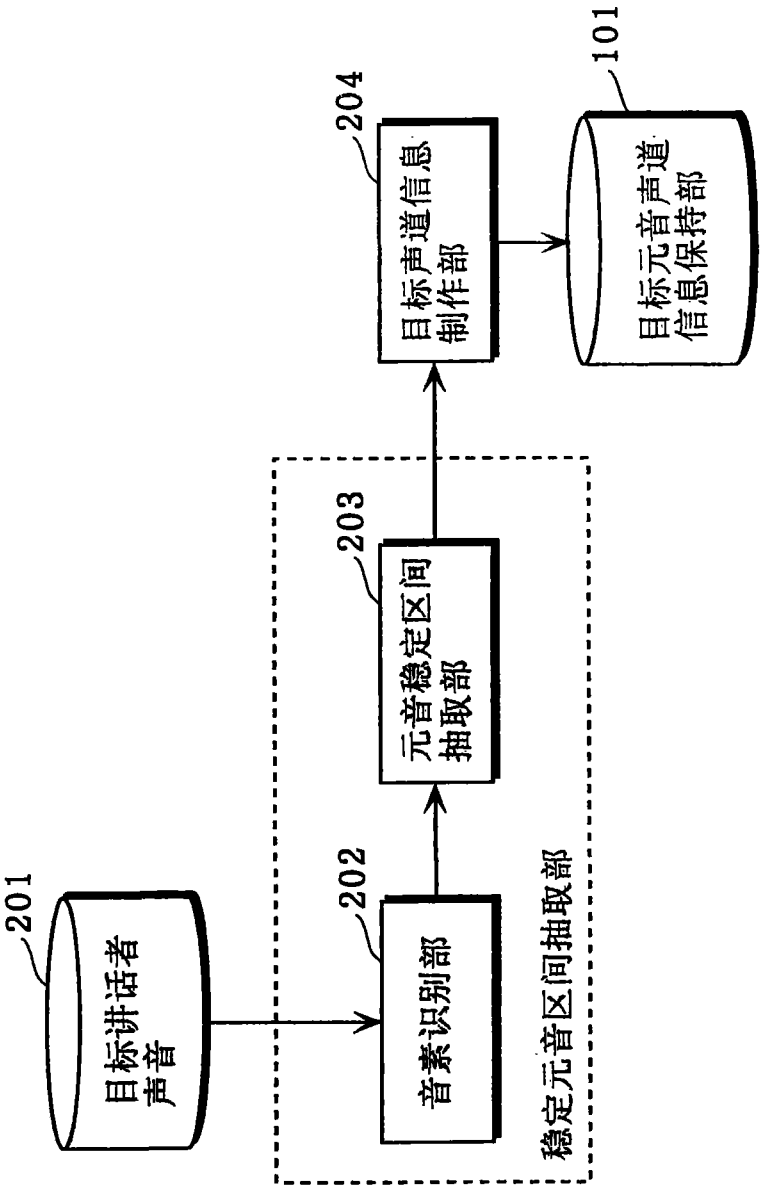

101
204
目标声道信息制作部
目标元音声道信息保持部
203
元音稳定区间抽取部
202
音素识别部
稳定元音区间抽取部
201
目标讲话者声音

图6

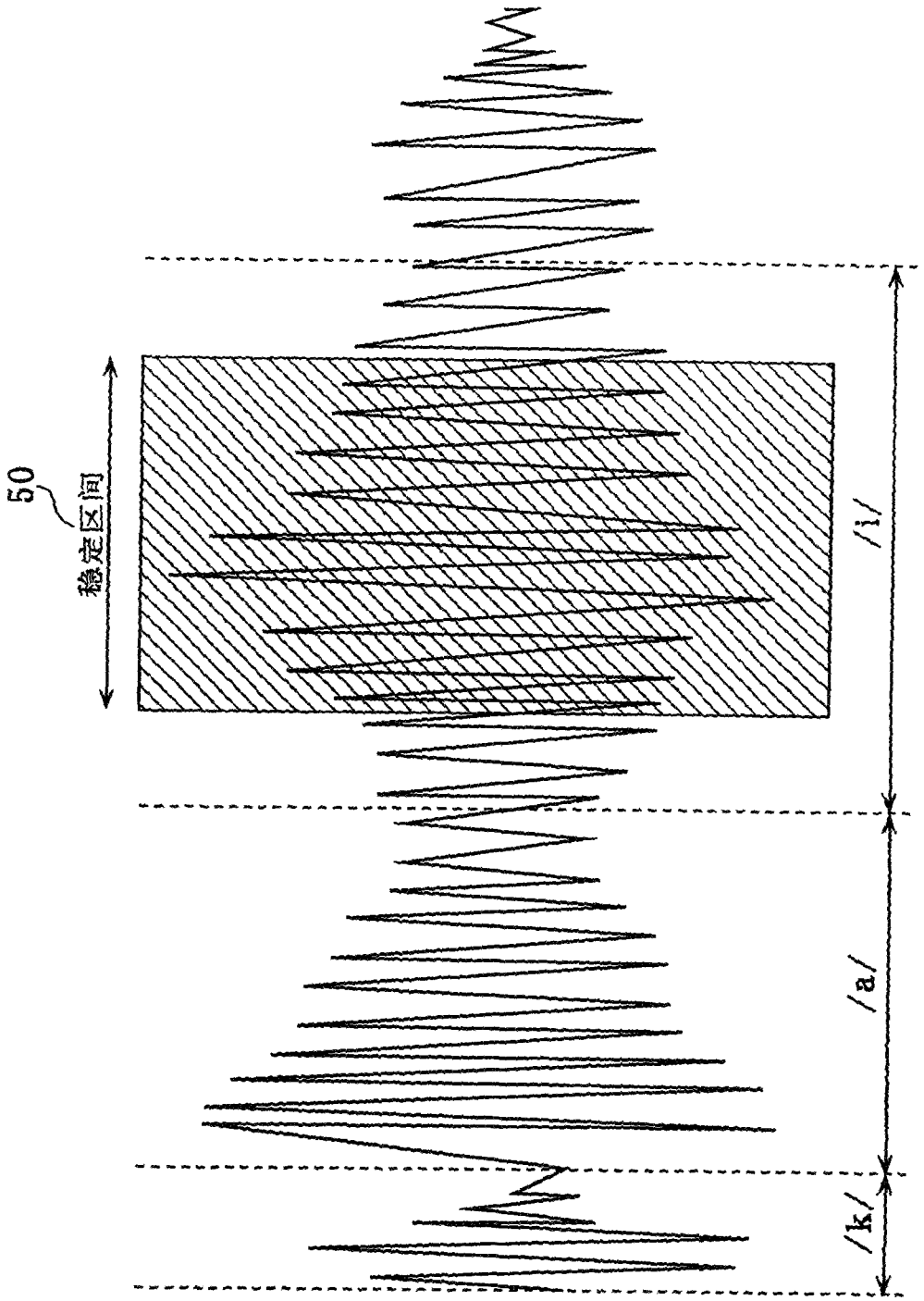


图 7

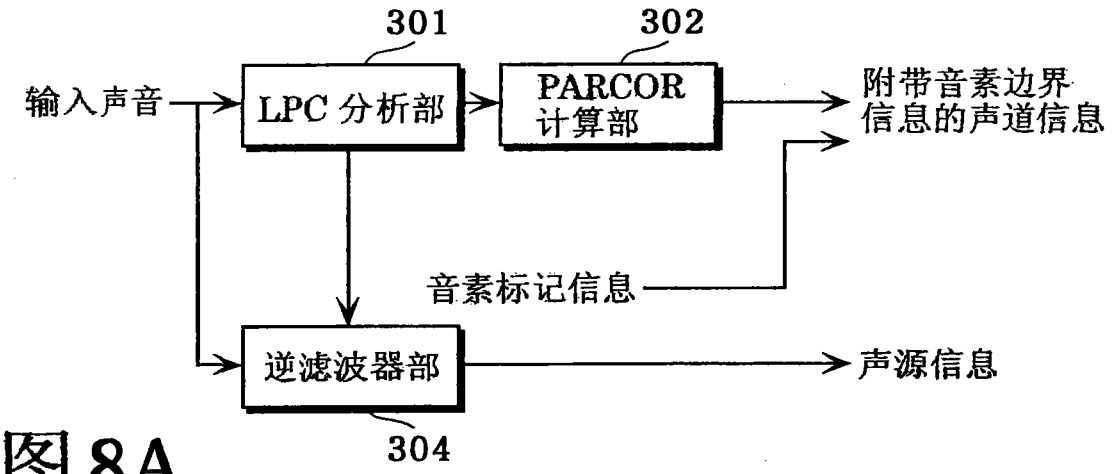

301
302
输入声音
LPC 分析部
PARCOR
计算部
附带音素边界
信息的声道信息
音素标记信息
逆滤波器部
声源信息
304

图 8A

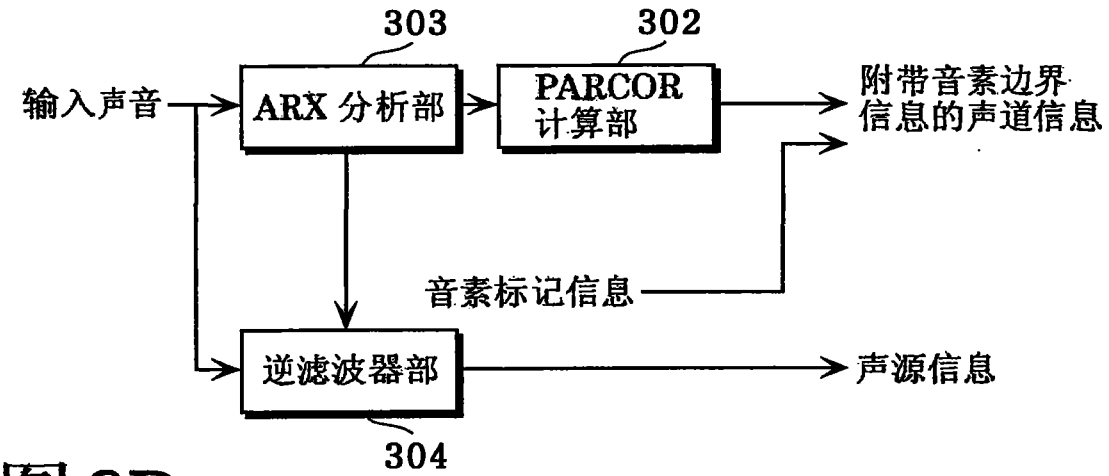

303
302
输入声音
ARX 分析部
PARCOR
计算部
附带音素边界
信息的声道信息
音素标记信息
逆滤波器部
声源信息
304

图 8B

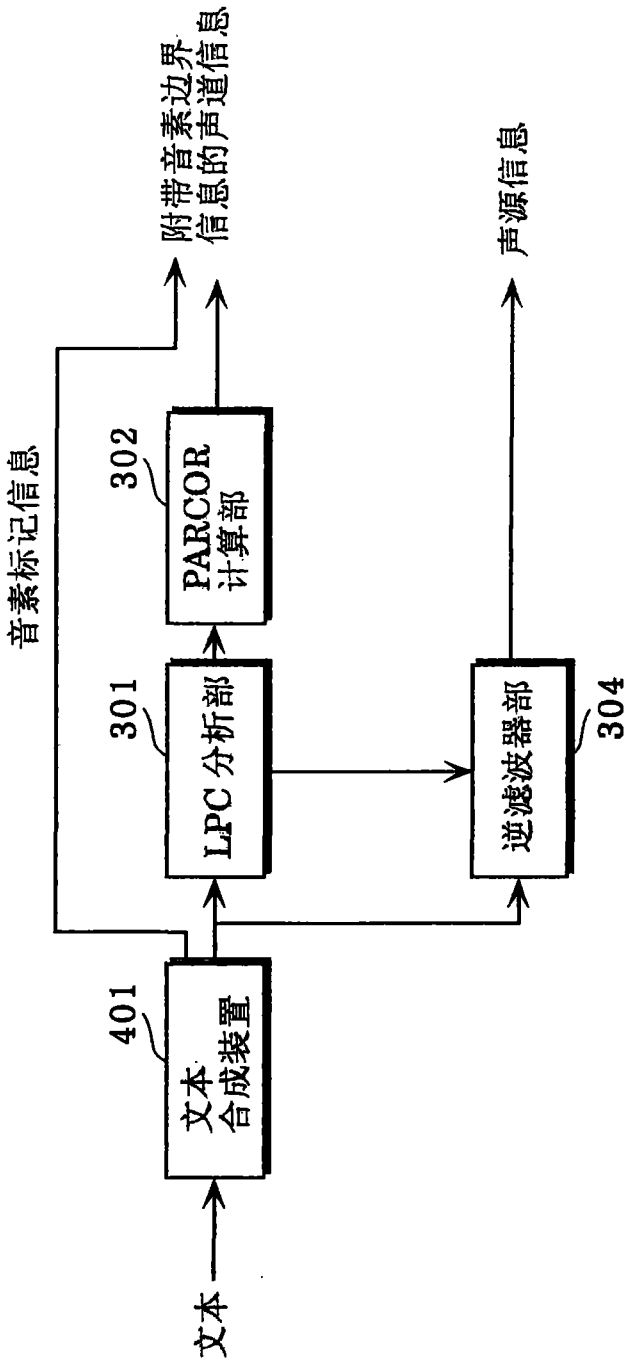

文本
401
文本合成装置
301
LPC 分析部
302
PARCOR 计算部
304
逆滤波器部
音素标记信息
附带音素边界信息的声道信息
声源信息

图 9

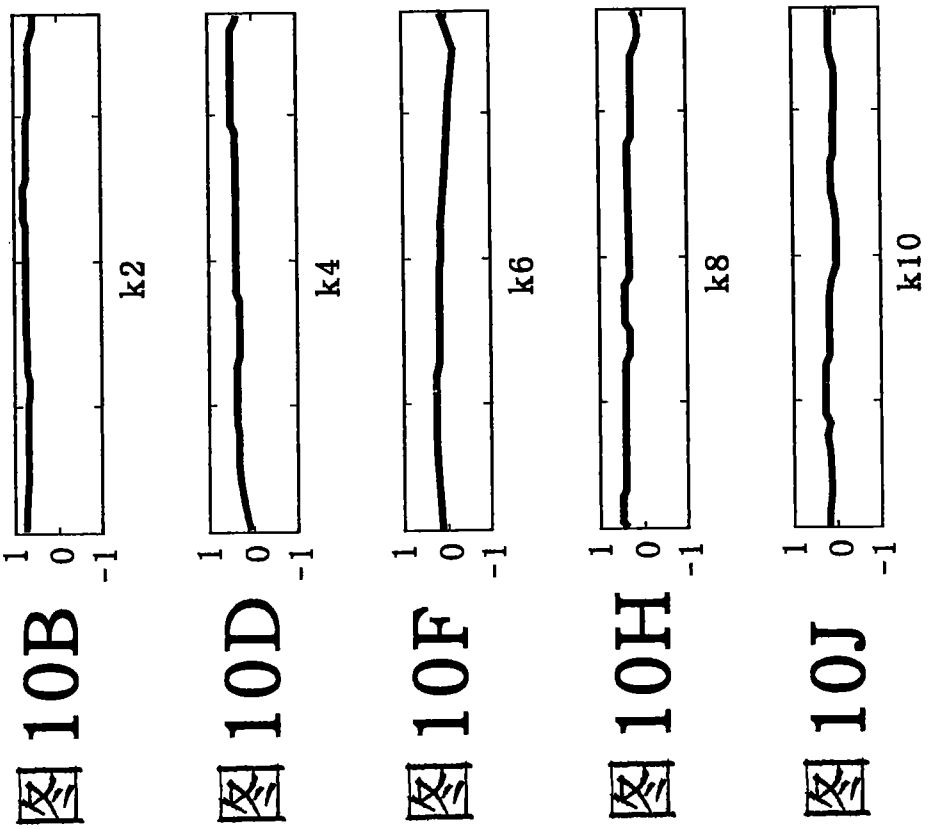


图10B 图10D 图10F 图10H 图10J

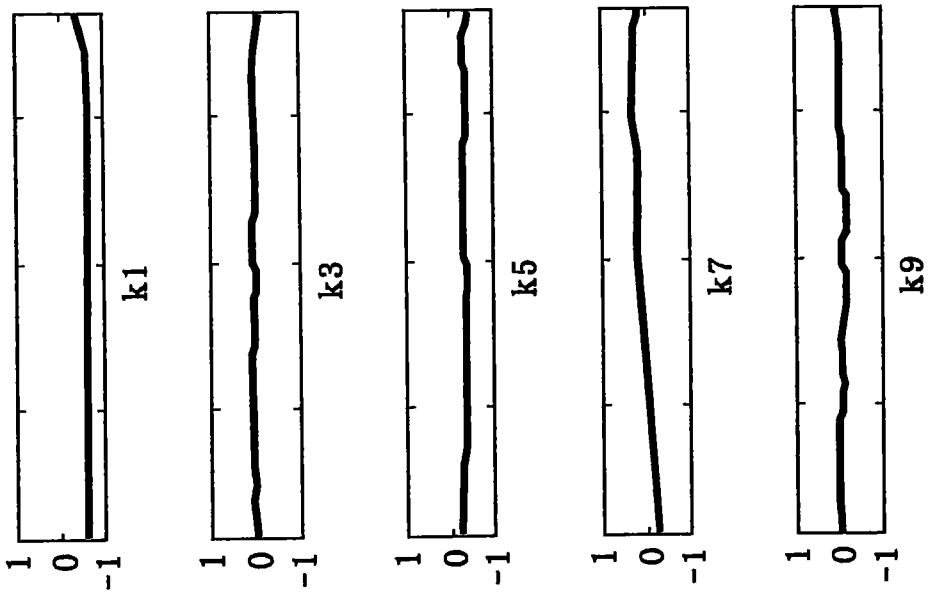


图10A 图10C 图10E 图10G 图10I

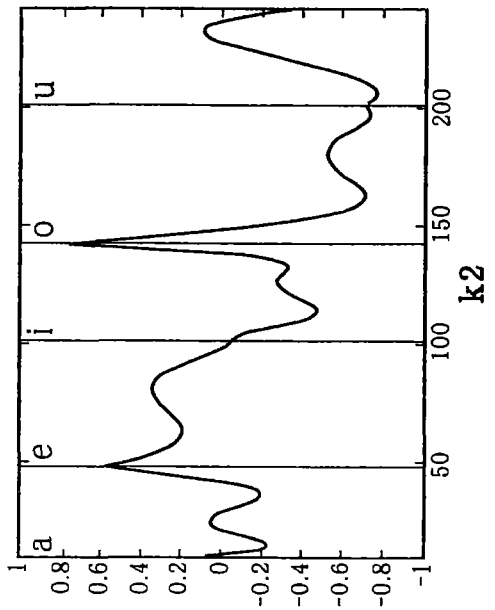


图11B

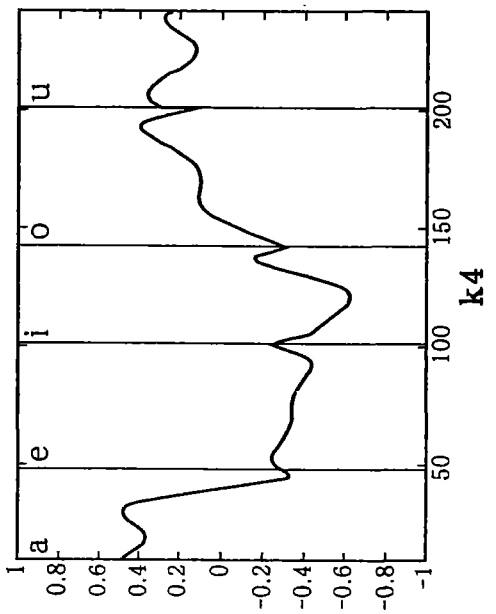


图11D

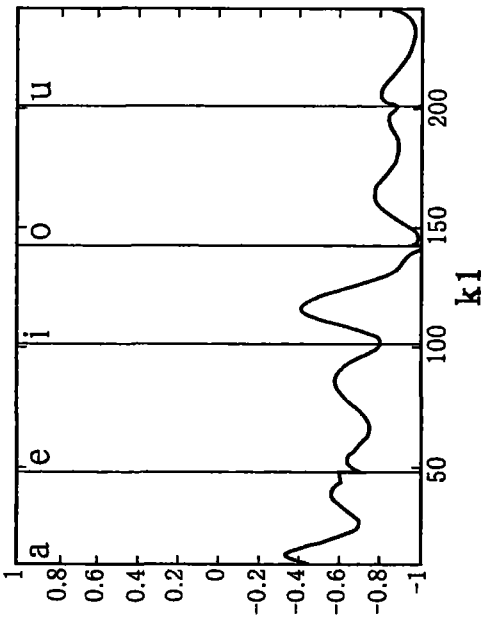


图11A

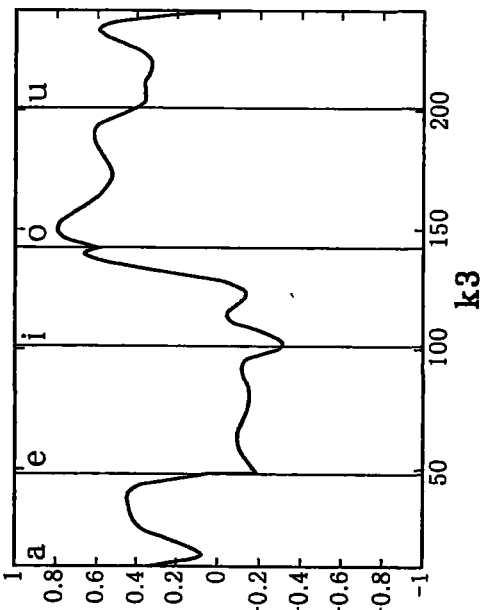
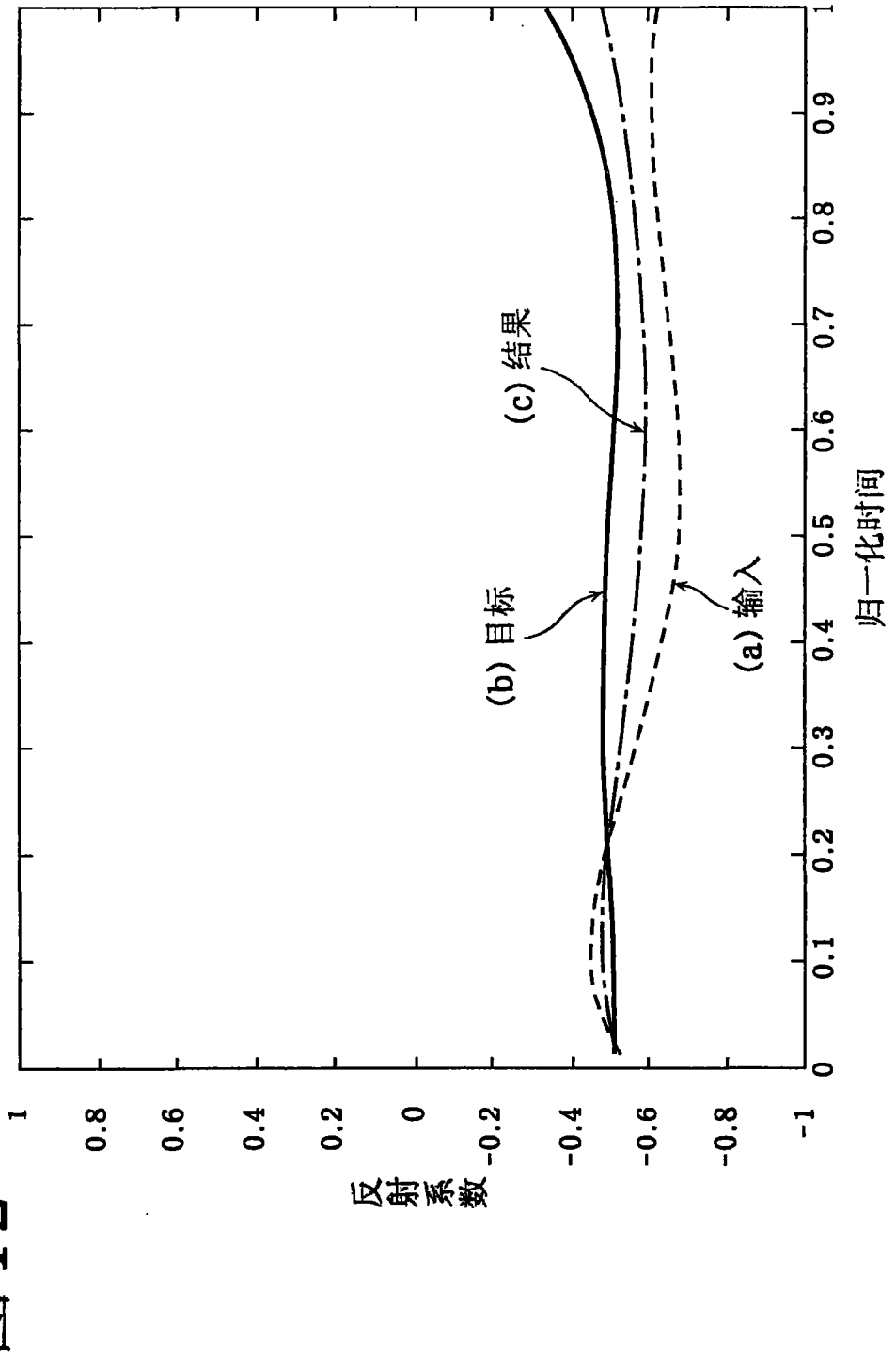


图11C

图12


反射系数
归一化时间
(a) 输入
(b) 目标
(c) 结果
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

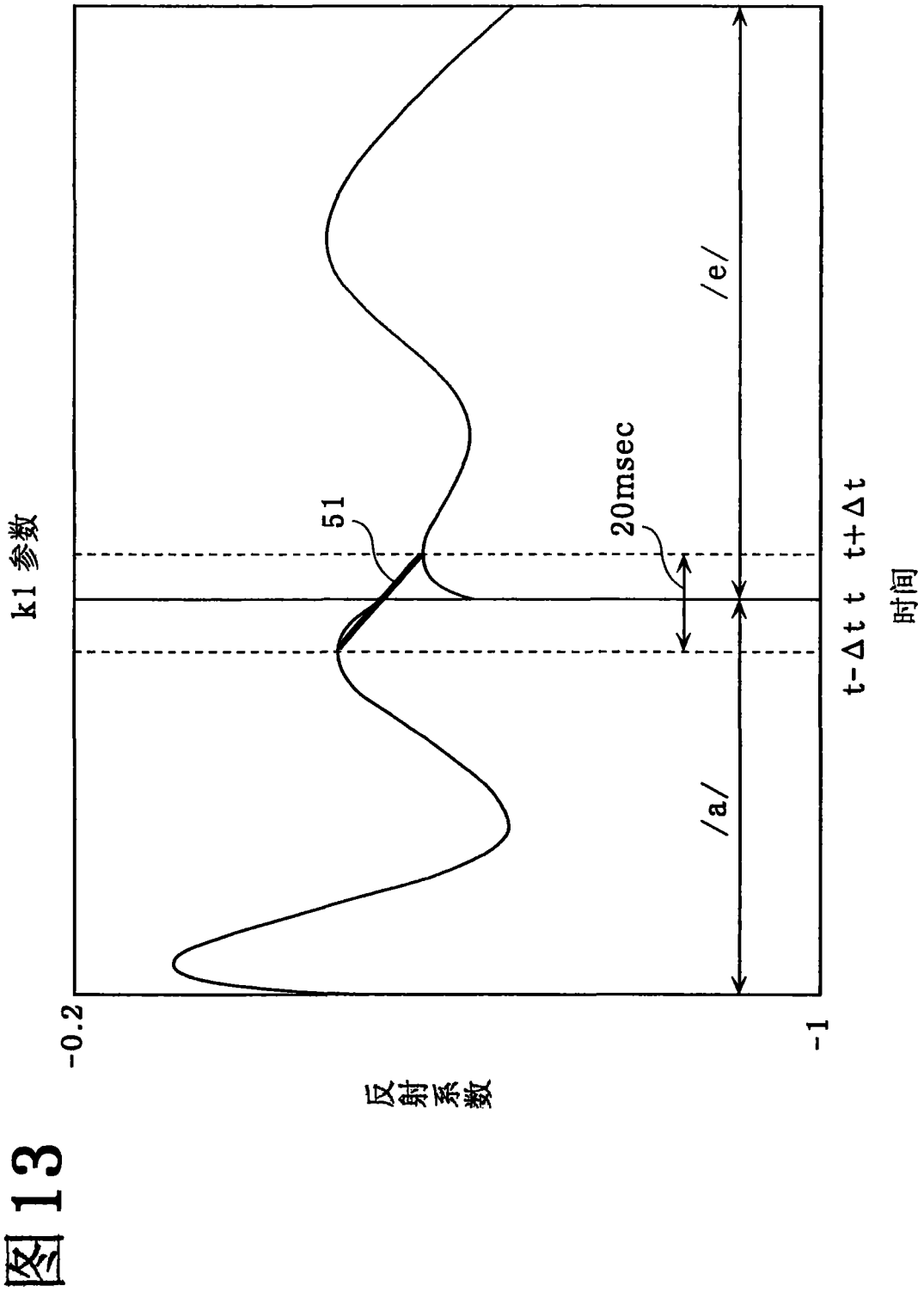


图 13

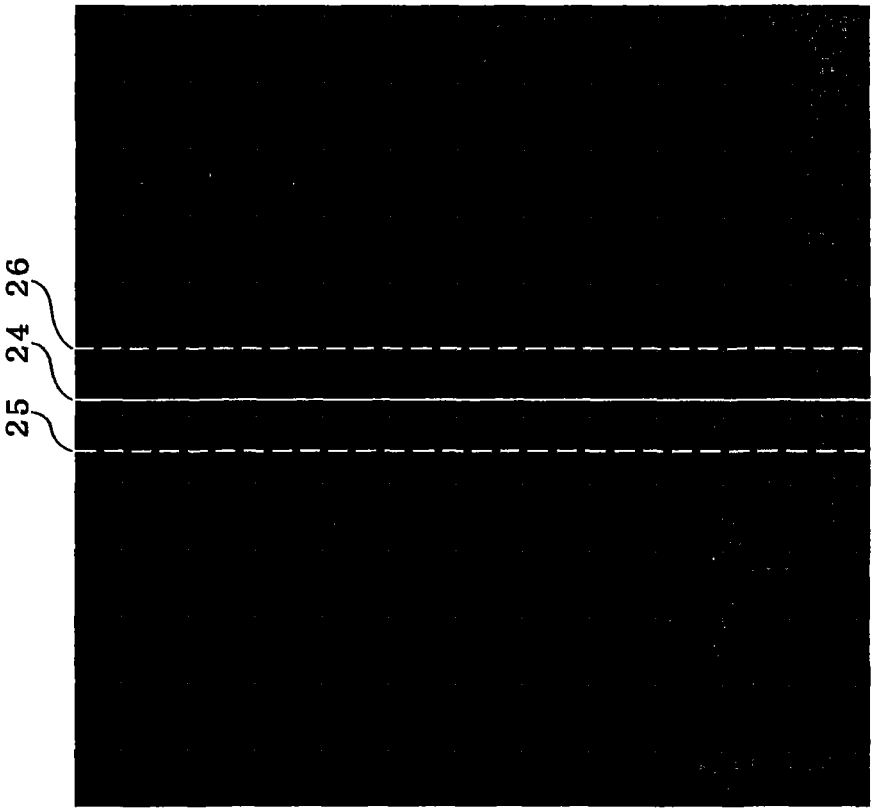


图14B

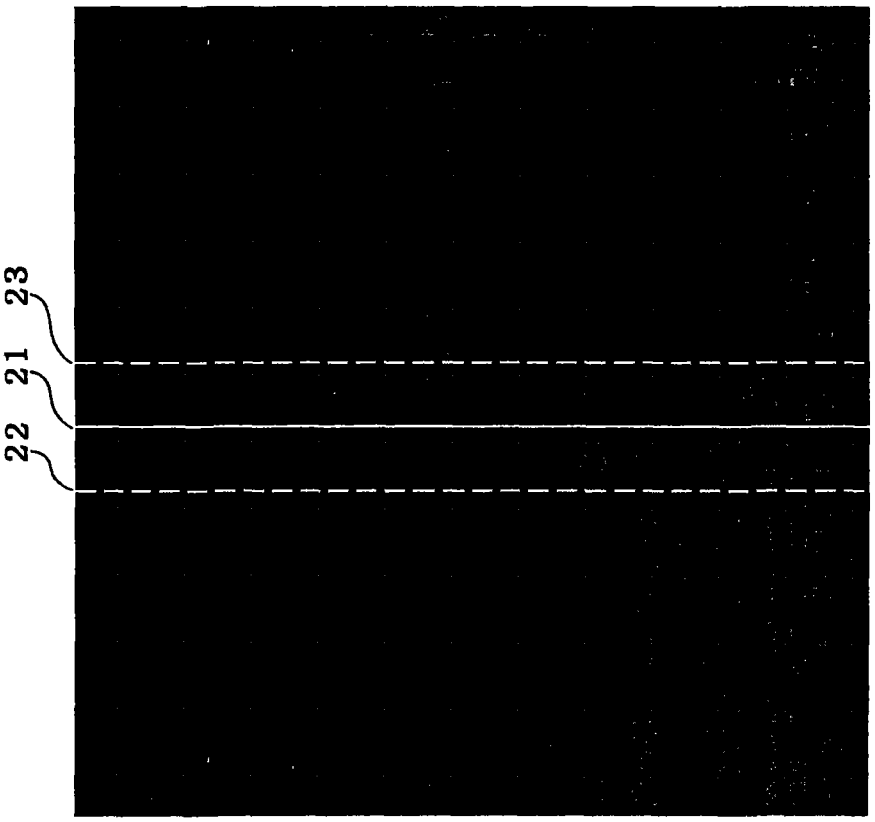


图14A

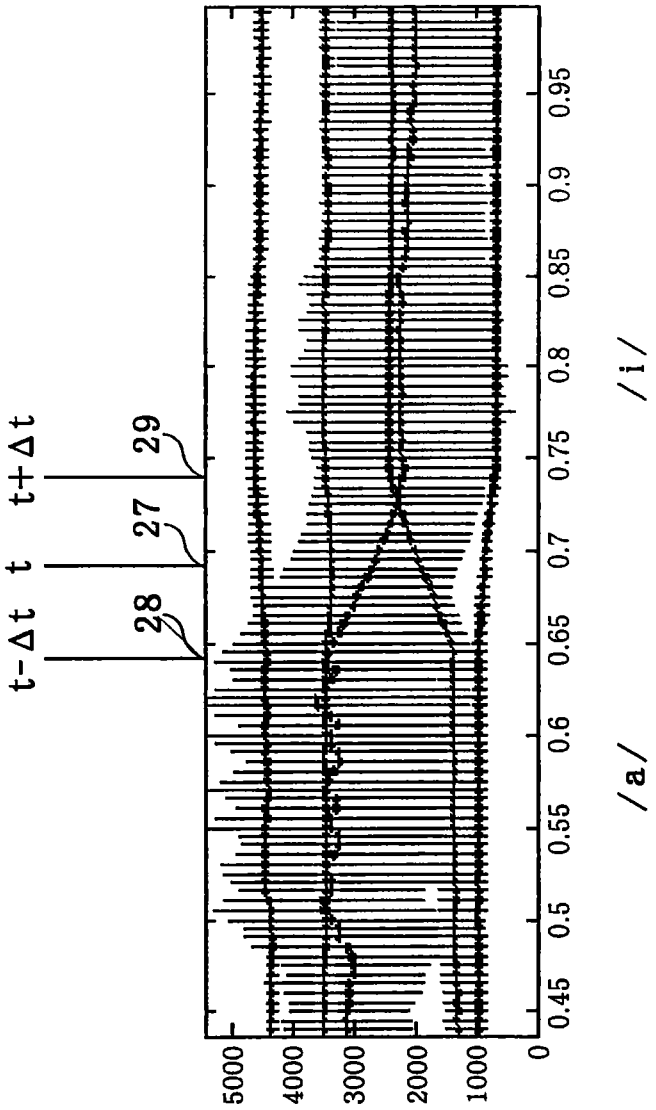

t-Δt t t+Δt
28
27
29
5000
4000
3000
2000
1000
0
0.45
0.5
0.55
0.6
0.65
0.7
0.75
0.8
0.85
0.9
0.95
/a/
/i/

图15

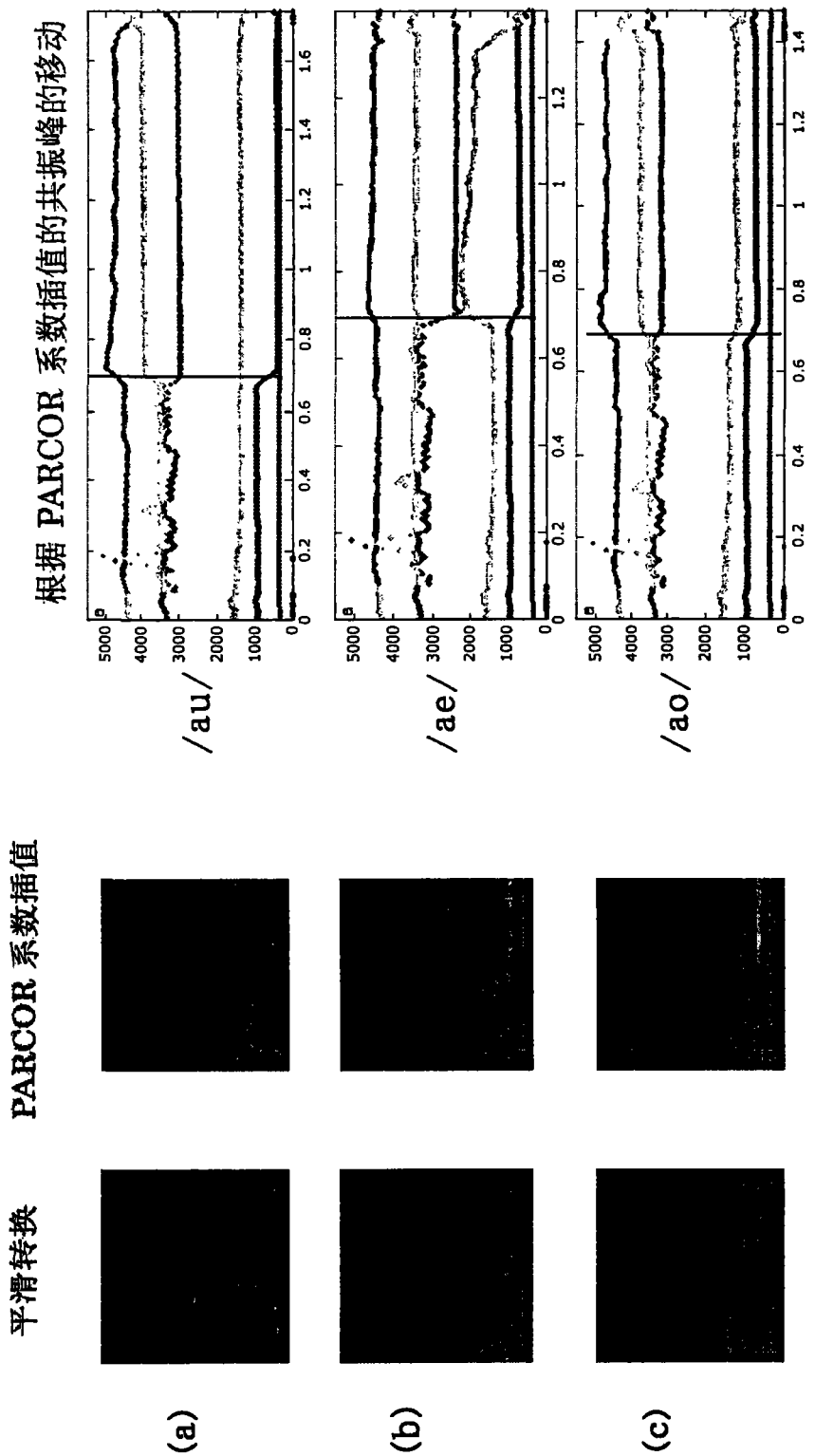
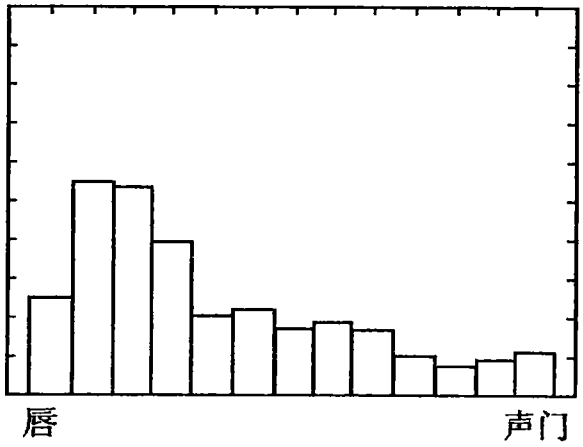


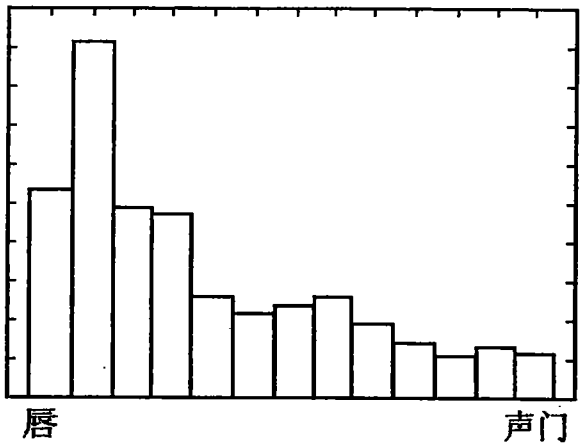
图 16

图17A



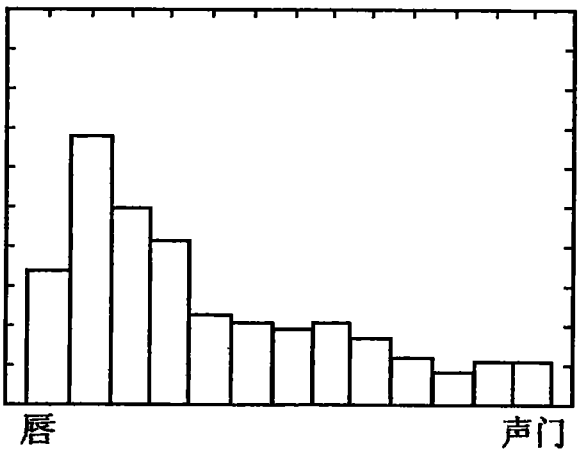
0%
（输入）

图17B



100%
（目标）

图17C



50%
（结果）

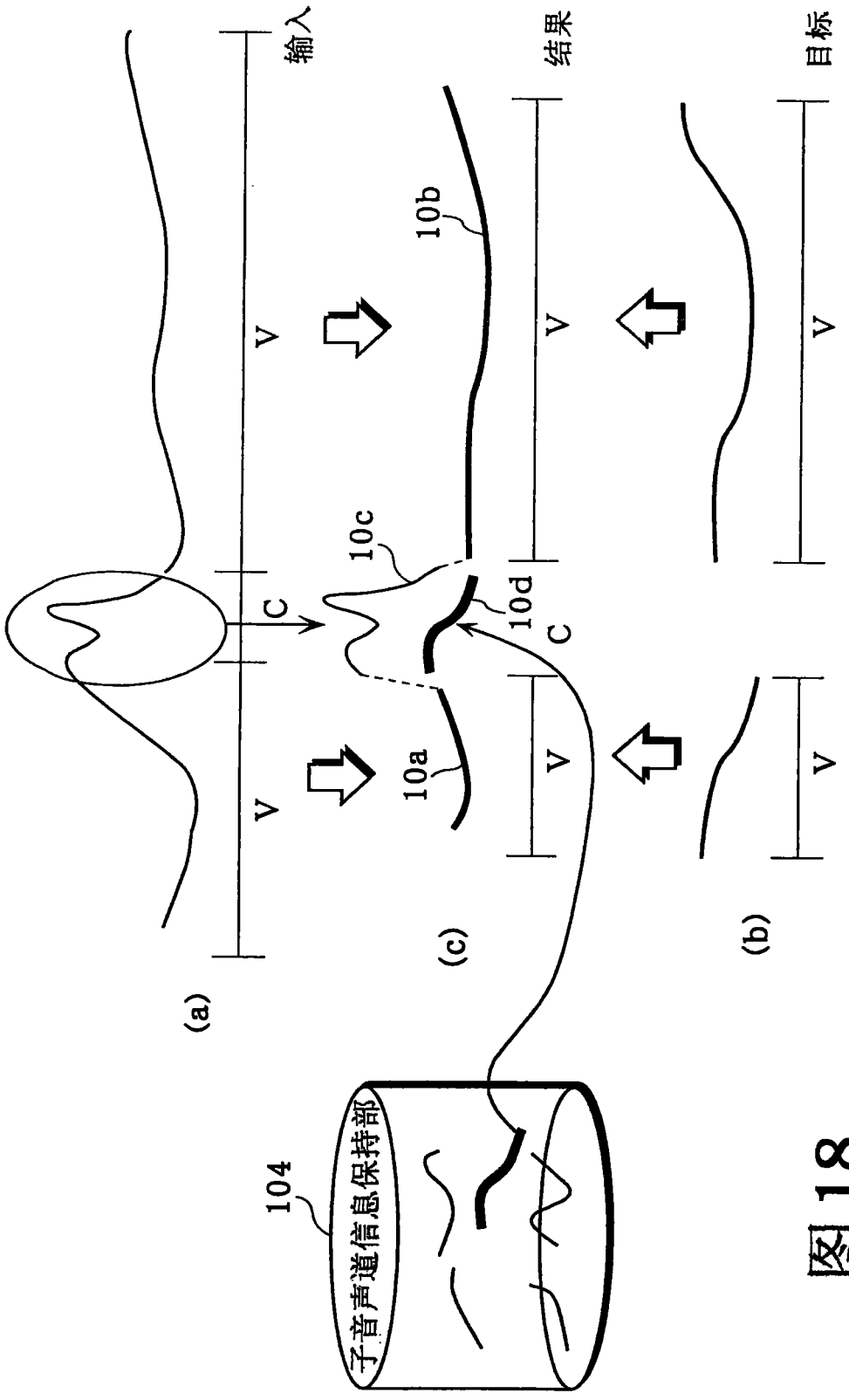


图 18

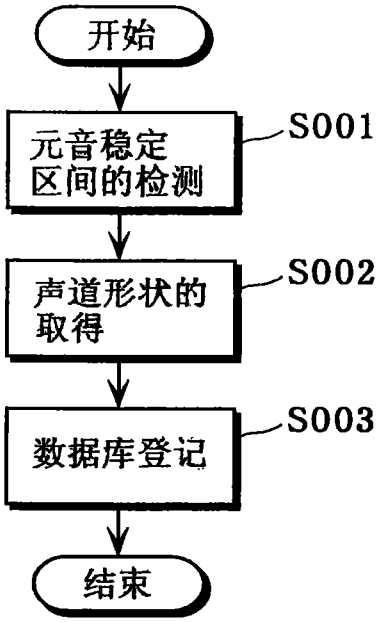

开始
元音稳定区间的检测
S001
声道形状的取得
S002
数据库登记
S003
结束

图19A

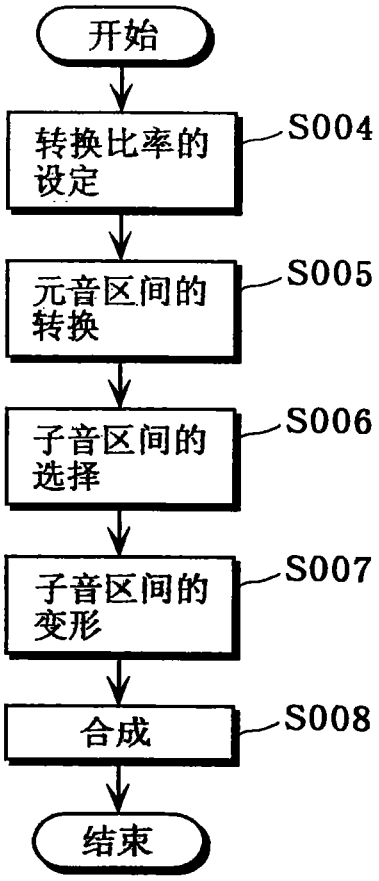

开始
转换比率的设定
S004
元音区间的转换
S005
子音区间的选择
S006
子音区间的变形
S007
合成
S008
结束

图19B

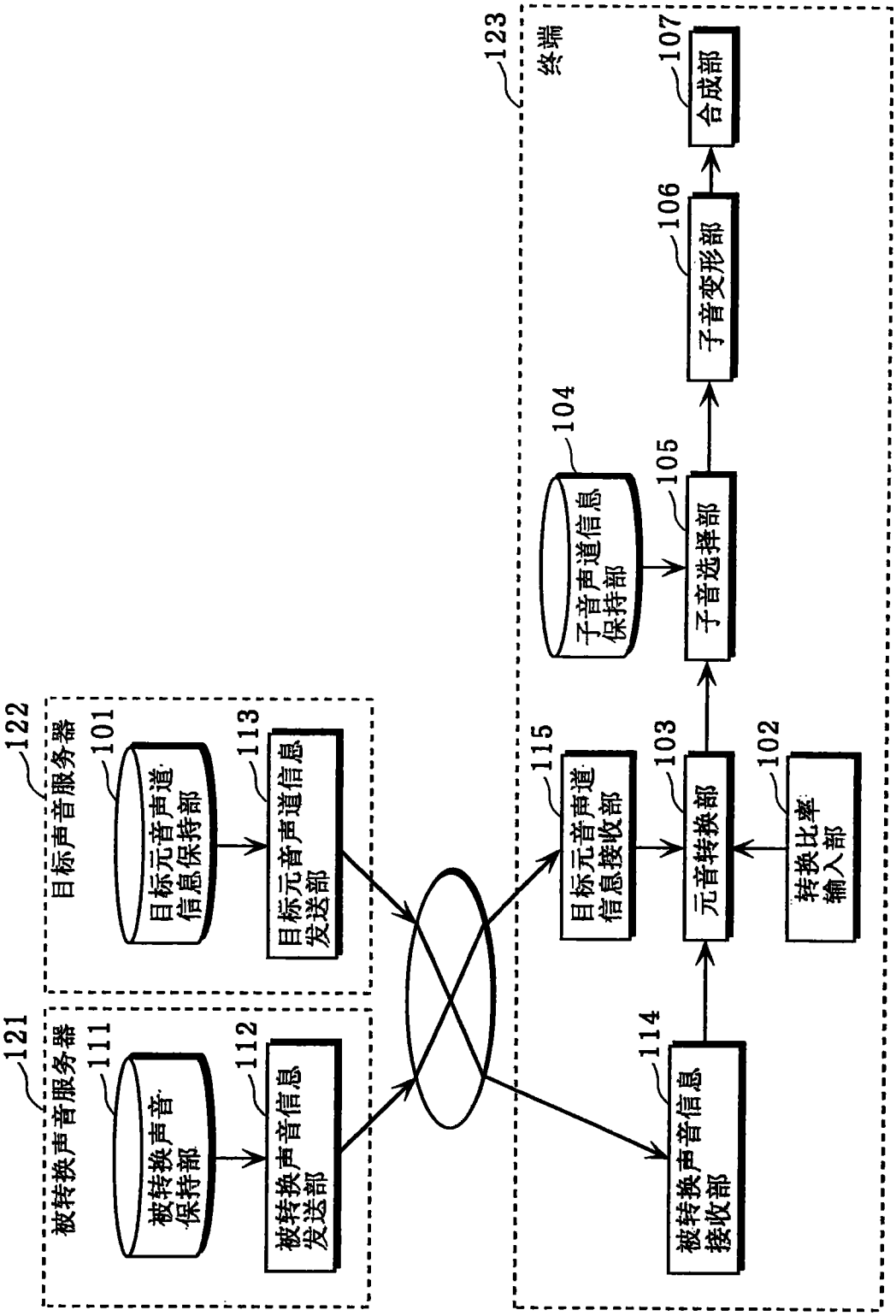

123
终端
107
合成部
106
子音变形部
104
子音声道信息保持部
105
子音选择部
115
目标元音声道信息接收部
103
元音转换部
102
转换比率输入部
114
被转换声音信息接收部
122
目标声音服务器
101
目标元音声道信息保持部
113
目标元音声道信息发送部
121
被转换声音服务器
111
被转换声音保持部
112
被转换声音信息发送部

图 20

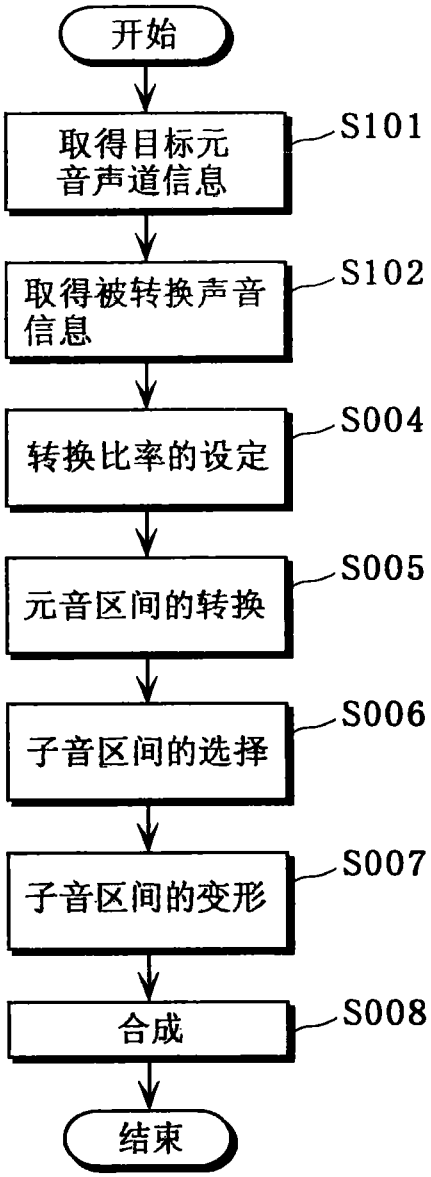

开始
取得目标元音声道信息
S101
取得被转换声音信息
S102
转换比率的设定
S004
元音区间的转换
S005
子音区间的选择
S006
子音区间的变形
S007
合成
S008
结束

图21

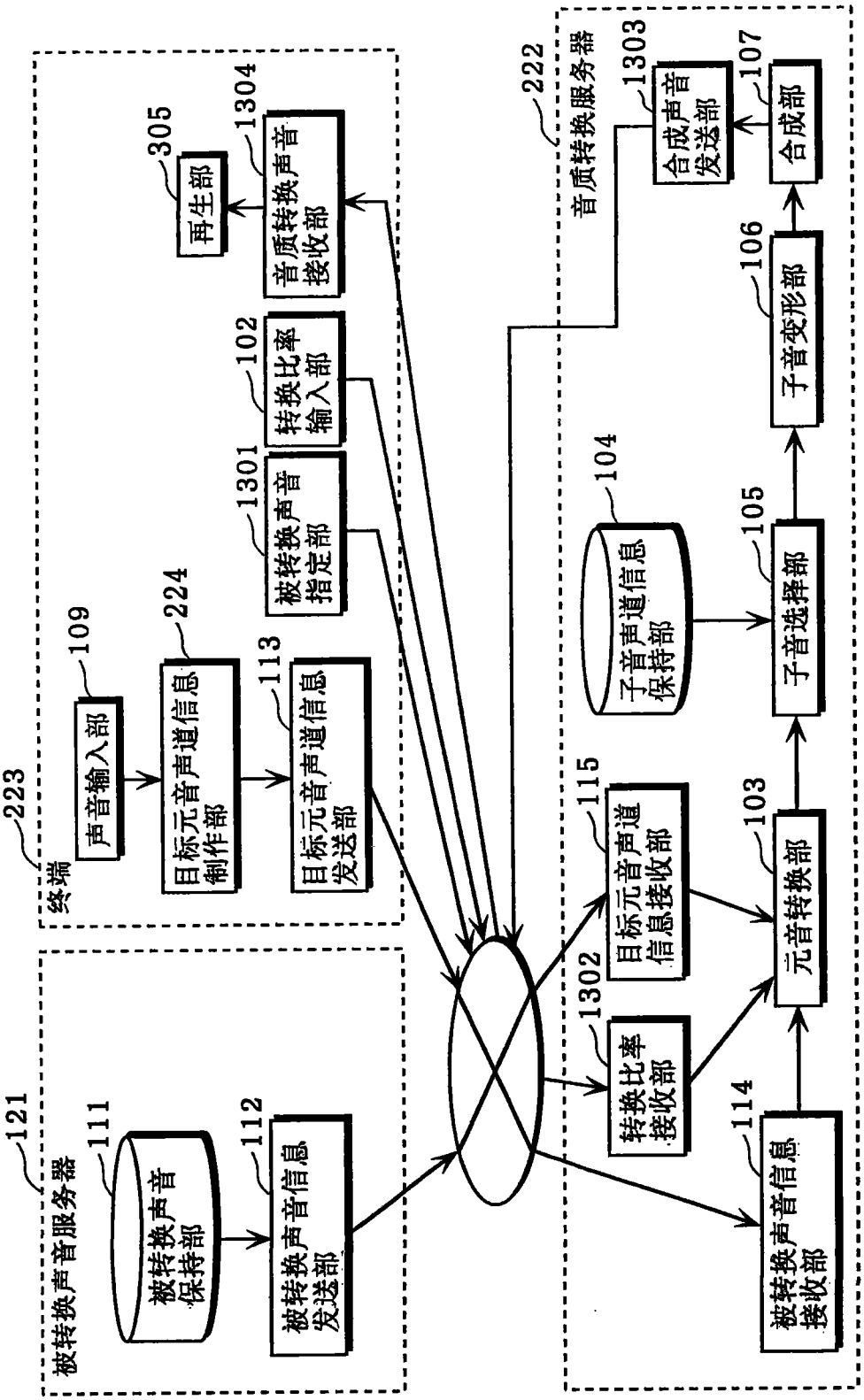

121
被转换声音服务器
111
被转换声音保持部
112
被转换声音信息发送部
223
终端
109
声音输入部
224
目标元音声道信息制作部
113
目标元音声道信息发送部
1301
被转换声音指定部
102
转换比率输入部
305
再生部
1304
音质转换声音接收部
222
音质转换服务器
1303
合成声音发送部
107
合成部
106
子音变形部
104
子音声道信息保持部
105
子音选择部
115
目标元音声道信息接收部
1302
转换比率接收部
103
元音转换部
114
被转换声音信息接收部

图 22

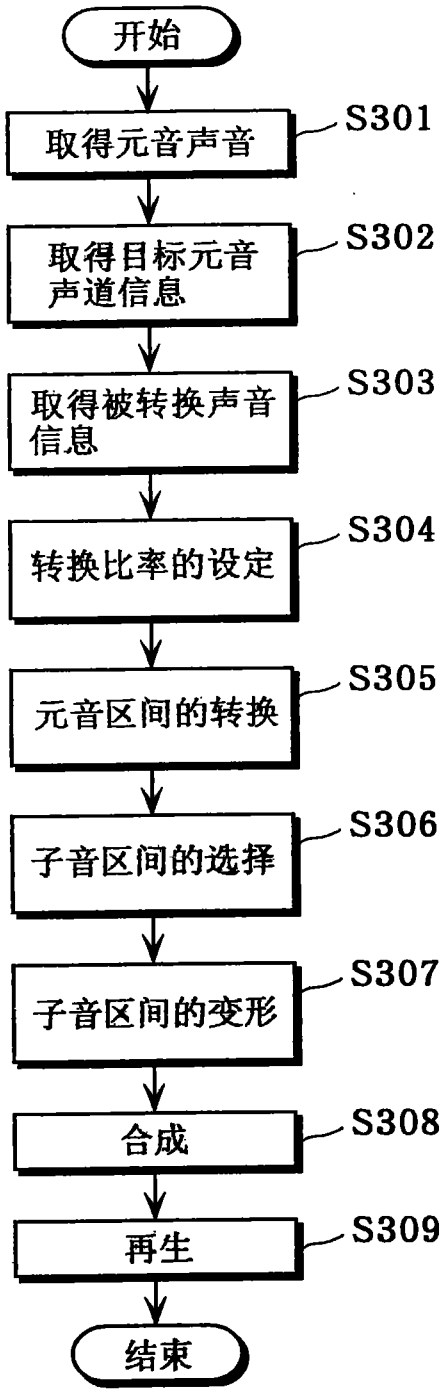

开始
取得元音声音
S301
取得目标元音声道信息
S302
取得被转换声音信息
S303
转换比率的设定
S304
元音区间的转换
S305
子音区间的选择
S306
子音区间的变形
S307
合成
S308
再生
S309
结束

图23