[19]中华人民共和国专利局

[51]Int.Cl[6]
G10L 3/00

[12] 发明专利申请公开说明书

[21] 申请号 97114182.7

[43]公开日 1998 年 8 月 12 日

[11] 公开号 CN 1190236A

[22]申请日 97.12.10

[30]优先权

[32]96.12.10[33]JP[31]329845/96

[71]申请人 松下电器产业株式会社

地址 日本国大阪府

[72]发明人 西村洋文 蓑轮利光 新居康彦

[74]专利代理机构 上海专利商标事务所

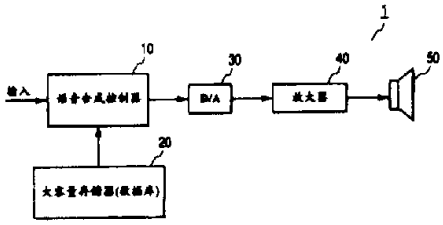
代理人 钱慰民

权利要求书 3 页 说明书 8 页 附图页数 15 页

[54]发明名称 语音合成系统及其减少冗余的波形数据库

[57]摘要

在一种语音合成系统中，将合成语音必须的一组音段的每个波形分成音调波形，把相似的音调波形归类到同一分组中。从每个分组中选择一个音调波形作为代表，并赋予其音调波形 ID。波形数据库至少包括一个音调波形指针表和一个音调波形表，前者的每条记录都包括音段 ID 和音调波形 ID，而后者包括音调波形 ID 和相应的音调波形。这缩小了波形数据库的大小。对于数据库缺少的音调波形，可以使用前表中一个与其相邻的音调波形 ID，不致使该音调波形畸变。



权 利 要 求 书

1. 一种在通过链接预定音段来合成语音的系统中使用的据库，其特征在于，所述数据库包括:

第一表，用于使每个所述预定音段与音调波形的音调波形识别码(ID)相联系，当按所述音调波形 ID 的列出的次序合并所述音调波形时，构成所述每个预定音段的波形；和

第二表，用于使每个音调波形 ID 与用所述每个音调波形 ID 识别的音调波形数据相联系。

2. 一种在通过链接预定音段来合成语音的系统中使用的数据库，每个所述音段由一个音素链接格式和一个音调带确定，其特征在于，所述数据库包括:

第一表装置，用于使每个所述预定音段与音调波形的音调波形 ID 相联系，其中所述预定音段由一个预定音调带 ID 和一个预定音素链接格式 ID 识别，当按所述音调波形 ID 的列出的次序合并所述音调波形时，所述音调波形构成所述每个预定音段的波形；和

第二表装置，允许用每个所述音调波形 ID 和所述一个预定音调带 ID 寻找与所述每个音调波形 ID 相关的音调波形数据。

3. 如权利要求 2 所述的数据库，其特征在于，所述第一表装置包括用音素链接格式构造的表，每个所述表中的每条记录都包含一个所述预定音调带 ID 和音调波形的音调波形 ID，当按所述音调波形 ID 的列出的次序合并所述音调波形时，所述音调波形构成一个波形，该波形由一个与所述每个表相关的音素链接格式和所述一个预定音调带 ID 所表征。

4. 如权利要求 2 所述的数据库，其特征在于，

所述第二表装置包括用音素分类的表分组，其中所述音素构成了由音素链接格式 ID 识别的音素链接格式;

每个所述表分组包括由所述预定音调带 ID 识别的表；和

每个所述表中的每条记录都包括由与所述每个表相关的一个音素链接格式和一个音调带确定的诸音调波形之音调波形 ID 中的一个音调波形 ID，以及一个与所述音调波形 ID 相关的音调波形。

5. 如权利要求 1 或 2 所述的数据库，其特征在于，所述数据库中的所有音调

波形数据都有相同的相位特征。

6. 一种在通过链接预定音段来合成语音的系统中使用的数据库，其特征在于，所述数据库包括:

第一表，用于使每个所述预定音段与音调波形和哑声波形的音调波形 ID 和哑声波形 ID 相联系，其中当按所述波形 ID 的列出次序合并时，音调波形和哑声波形构成所述每个预定音段的波形;

第二表，用于使每个哑声波形 ID 与用所述每个哑声波形 ID 识别的哑声波形数据相联系，其中包含接近相似的哑声波形的音段具有与所述第一表中分配给所述接近相似之哑声波形相同的波形 ID 。

7. 一种用于制作在通过链接预定音段来合成语音的系统中使用的数据库的方法，其特征在于，所述方法包括以下步骤:

将每个所述预定音段分成音调波形;

将所有音调波形归类到由非常相似的音调波形组成的分组中;

在每个所述分组中选择一个所述非常相似的音调波形;

为每个所述分组的所述被选中音调波形分配一个音调波形 ID ;

产生第一表，对于每个所述分组，表中有一条记录，所述记录包含所述音调波形 ID 和所述所选的音调波形的数据; 并

产生第二表，表记录 ID 包括所述预定音段的 ID ，所述第二表的每条记录包括音调波形 ID ，当按所述音调波形 ID 列出次序合并时，所述音调波形构成由所述记录 ID 识别的波形。

8.如权利要求 7 所述的方法，其特征在于，将所有音调波形归类的所述方法包括用每个所述音调波形的频谱参数将所有音调波形归类的步骤。

9. 如权利要求 7 所述的方法，其特征在于，在每个所述分组中选择一个所述非常相似的音调波形的所述步骤包括在每个所述分组中选择一个功率最大的音调波形的步骤。

10. 如权利要求 7 所述的方法，其特征在于，如此实现在每个所述分组中选择一个所述非常相似的音调波形的所述步骤，使得所有被选中的音调波形具有相同的相位特征。

11.一种通过链接预定音段来合成语音的系统，其特征在于，包括:

用于从所述预定音段中确定对于所述语音必须的音段之 ID 的装置;

用于将每个所述确定的 ID 与音调波形 ID 相联系的装置，其中当按所述音调波形 ID 的列出次序合并时，所述音调波形 ID 的音调波形构成一个由所述每个经确定的 ID 识别的波形；

用于获取与所述音调波形 ID 相关的音调波形的装置；

用于合并所述获得的音调波形以形成所述必要音段的装置；

用于合并所述必要音段以产生所述语音的装置。

12. 一种通过链接预定音段来合成语音的系统，每个预定音段由一个音素链接格式和一个音调带确定，其特征在于，所述系统包括：

用于从所述预定音段中确定对于所述语音必须的音段之 ID 的装置；

用于将每个所述确定的 ID 与音调波形 ID 相联系的装置，其中当按所述音调波形 ID 的列出次序合并时，所述音调波形 ID 的音调波形构成一个由所述每个被确定的 ID 识别的波形；

用于获取与所述音调波形 ID 相关的音调波形的装置；

用于合并所述获得的音调波形以形成所述必要音段的装置；

用于合并所述必要音段以产生所述语音的装置。

说　明　书

语音合成系统及其减少冗余的波形数据库

本发明涉及一种用相当小的波形数据库提供更自然合成语音的语音合成系统和方法。

在使用某种语言的常规语音合成系统中，每个语音都被分成长度比该语言中所用单词长度更短的音段(音素链组元或合成单元)。需形成并存储用该语言进行语音合成所必须的一组如此音段的波形数据库。在合成过程中，将所给的文本分成音段，并借助波形数据库将与所分音段相关的波形合成所给文本相应的语音。1996 年日本未审查专利公报 Hei8-234793 揭示了一种这样的语音合成系统。

但在常规的系统中，如果某个音段与存储在数据库中的任何音段都不同，那么即使数据库中存在一个或多个音段，其波形的之大部分与该音段波形相同，则该音段也会作为一个不同的音段被存储在数据库中，这使得数据库非常冗长．如果为了避免冗长而限制数据库中的音段数目，那么在语音合成过程中，必然会因每次缺少音段而使任何受限的音段畸变，造成合成语音质量劣化。

本发明的一个目的是提供一种语音合成系统和方法，允许将波形数据库的规模做得较小，并对于波形数据库中缺少的音段避免任何音段的畸变，从而提供令人满意的语音合成质量。

用一种系统可以实现上述目的，在该系统中，进一步将对应于某种语言中标准音段(音素链组元)的每个波形分成音调波形，并把密切相似的音调波形归成一组。从每个分组的音调波形中选择一个作为该分组的代表，并给予其一个音调波形 ID．波形数据库至少包括一个(音调波形指针)表和一个(音调波形)表，前表中的每条记录都包括每个音段的音段 ID 和音调波形 ID，当按所列顺序合并时，其音调波形构成一个由音段 ID 识别的波形，而后表包括音调波形 ID 和相应的音调波形。这使得不同但相似的音段共享公用的音调波形，缩小了波形数据库的大小。对于数据库缺少的每个音调波形，使用与该缺少音调波形最相似的音调波形，也就是说，使用音调波形指针表中与该缺少的音调波形 ID 相邻的音调波形 ID 中的一个，而不引起音调波形的畸变。

结合附图阅读以下关于本发明较佳实施例的描述，将更清楚本发明的其它目

的和优点，其中：

图 1 是一示意方框图，示出了一例实施本发明原理的语音合成系统；

图 2 是一示意图，例示了如何根据基于 VCV 的语音合成方案合成日语单词‘inu’和‘iwashi’；

图 3 是一流程图，示出了依照本发明一说明性实施例形成有声发音波形数据库的过程；

图 4A 是一示意图，例示了在图 3 步骤 350 中形成的音调波形指针表；

图 4B 是一示意图，例示了在图 3 步骤 340 中建立的音调波形表的每条记录的结构；

图 5A 和 5B 是流程图，分别例示了获得周期性波形和音调波形之频谱包络线的过程；

图 6 是一曲线图，示出了一周期性波形的功率频谱；

图 7 是一示意图，示出了第一例在图 3 之步骤 330 中从一分组音调波形中选取一代表性音调波形的方法；

图 8 是一示意图，示出了第二例在图 3 之步骤 330 中从一分组音调波形中选取一代表性音调波形的方法；

图 9 是一示意图，示出了依照本发明第二说明性实施例在图 1 语音合成系统中使用的波形数据库的结构；

图 10 例示了图 9 中所示音调波形指针表，例如(有关音素链格式‘inu’的)9606inu 的结构；

图 11 是一流程图，示出了形成图 9 中有声发音波形数据库 900 的过程；

图 12 是一示意图，示出了不同音段是如何共享一个公用哑声的；

图 13 是一流程图，示出了依照本发明说明性实施例形成哑声(voiceless sound)波形数据库的过程；

图 14 是一流程图，例示了使用图 4 之有声发音波形数据库的语音合成程序的流程；以及

图 15 是一流程图，例示了使用图 9 和图 10 之有声发音波形数据库的语音合成程序的流程。

在所有附图中，相同的标号表示不止一个附图中使用的相同的单元。

图 1 中的语音合成系统包括依照本发明原理工作的语音合成控制器 10；用

于存储控制器 10 工作中使用的波形数据库的大容量存储器 20；用于将合成数字语音信号转换成模拟语音信号的数字-模拟转换器 30；以及用于提供合成语音输出的扬声器 50．大容量存储器 20 可以是存储量足够大的任何类型的存储器，例如可以是硬盘、 CD-ROM(光盘只读存储器)等等．如本领域中公知的，语音合成控制器 10 可以是包括图中未显示的 CPU(中央处理器，例如市场上可买到的微处理器)、图中未显示的 ROM(只读存储器)、未显示的 RAM(随机存取存储器)和接口电路(未显示)的任何合适的常规计算机．

尽管通常将以下描述的依照本发明原理的波形数据库存储在比 IC 存储器便宜的大容量存储器 20 中，但也可将其存储在控制器 10 的未显示的 ROM 中．可以将依照本发明原理进行语音合成的程序存储在控制器的未显示的 ROM 中，也可将其存储在大容量存储器 20 中．

波形数据库

以下结合常规语音合成方法描述说明性实施例，其中常规是将诸如 CV(C 和 V 分别是‘辅音’和‘元音’的缩写)、 VCV 、 CV/VC 或 CV/VCV 等链接波形衔接起来，合成语音．具体地说，假设如图 2 所示，以下实施例基本上将 VCV 链接波形用作语音的音段或语音成分组成，例如该图示出了如何依照基于 VCV 的语音合成方法来合成日语单词‘inu’和‘iwashi’．在图 2 中，通过合并语音成分或音段 101 至 103 合成单词‘inu’．通过合并音段 104 至 107 合成单词‘iwashi’．语音成分 102 、 105 和 106 是 VCV 成分，语音成分 101 和 104 是单词开始的成分， 而语音成分 103 和 107 是单词结尾的成分．

图 3 是一流程图，示出了依照本发明一说明性实施例形成有声发音波形数据库的过程．在图 3 中，首先在步骤 300 准备一组进行日语语音合成必须的音段抽样．为此，对包含这些音段的各种单词和语音进行实际发音，并将其存储在存储器中．将存储的音素波形分成基于 VCV 的音段，从中选出必要的音段，并将其集中到一个图中未示出的音段表(即，音段抽样组)中，表中的每条记录包括一个音段识别码(ID)及其相应的音段波形．

在步骤 310 中，将音段表(未示出)中的每个音段波形再分成如图 2 所示的音调波形．在该情况下，如果将每个音段细分成音素或语音单元，分割单元还不够小，不容易在分割的音素中找出相似的音素．例如，如果将 VCV 音段‘ama’

分成‘a’、‘m’和‘a’，那么不可能认为前后元音‘a’的发音是相同的，这无助于缩小波形数据库的大小．前元音‘a’与单个‘a’相似，而接下来的辅音‘m’对后元音‘a’有很大影响。所以，在图 2 中，分别将 VCV 音段 102 和 106 细分成音调波形 110 至 119 和波形 120 至 129 。通过这般处理，可以在细分的音调波形中找出许多接近相似的音调波形.在图 2 的情况下，音调波形 110 、111 与 120 之间非常相似.

在步骤 320 中，将细分的音调波形归类到由接近相似的音调波形组成的分组中。在步骤 330 中，用以下所述的方式从每组中选择一个音调波形作为代表，并为该选中的音调波形或分组指定一个音调波形 ID ，以便用该选中的音调波形代替该分组中的其它音调波形。在步骤 340 中，产生一个音调波形表，该表中的每条记录都包含一个被选的音调波形 ID 和该选择的音调波形的数据表示，完成有声发音的波形数据库．然后，在步骤 350 中，产生一个音调波形指针表，在该表中，抽样组中每个音段的 ID 与构成该音段之音调波形所属的分组中的音调波形 ID 相关联．哑声的波形数据库可以用常规的方法来形成。

如上所述，在音段间共享共同(非常相似的)音调波形可以大大缩小波形数据库的大小。

图 4A 例示了在图 3 之步骤 350 中形成的音调波形指针表。在图 4A 中，音调波形指针表 360 包括一个音段 ID 字段、多个音调波形 ID 和标志信息。音调波形 ID 字段包含这样一些音调波形的 ID，这些音调波形构成了由音调波形 ID 识别的音段。如果在表 360 的某条记录中存在属于同一音调波形分组的音调波形，那么这些音调波形的 ID 将是相同的。标志信息字段包含音段前元音的音调波形数，辅音的音调波形数，以及音段后元音的音调波形数。

图 4B 例示了在图 3 之步骤 340 中产生的音调波形表每条记录的结构．如图 4B 所示，音调波形表的每条记录包含一音调波形 ID 以及相应的音调波形数据。

以下将描述在图 3 之步骤 320 中把音调波形归类到音调波形接近相似的分组中的方法。具体地说，将讨论用例如音调波形之功率频谱和线性预测编码(LPC)频谱等频谱参数进行分类。

为了获得周期性波形的频谱包络线，必须遵循图 5A 所示的过程。在图 5A 中，步骤 370 中，对周期性波形进行傅里叶变换，产生如图 6 中 501 所示的对数功率频谱。然后，在步骤 380 对获得的频谱再作一次傅里叶变换，在步骤 390 进

行提升(liftering)并在步骤 400 进行傅里叶逆变换，最终产生如图 6 中 502 所示的频谱包络线。另一方面， 对于音调波形的情况，通过在步骤 450 将音调波形傅里叶变换成对数功率频谱来获得音调波形的频谱包络线。考虑到这一点，便不必如以前那样通过几十毫秒大小的分析窗来分析语音波形，而是在细分成音调波形后计算功率频谱。通过把功率包络线作为分类标准对音素进行分类，便可用少量的计算获得正确的分类。

图 7 是一示意图，例示了在图 3 之步骤 330 中从分类组的音调波形中选择一代表性音调波形的第一种方法。在图 7 中，标号 601 至 604 表示合成单元或音段。图中还用波形 605 的形式详细显示了音段 604 的后半段，其中波形 605 被细分成音调波形。将从波形 605 中截取的音调波形分成两组，即包括音调波形 611 和 612 的组 610，和包括功率频谱相似的音调波形 621 至 625 的组 620。最好从每组分组 610 和 620 中选择幅值最大的音调波形(611， 621)作为代表，以免在用所选音调波形替代诸如 621 等较大音调波形时造成信噪比降低。为此，在组 610 中选择音调波形 611，而在组 620 中选择音调波形 621。用这种方法选择代表性音调波形可以提高波形数据库的总的信噪比。由于一个音调波形组中自然会存在以不同音段截取的音调波形，所以即使在抽样组准备过程中记录了信噪比较低的音段，仍可能由从其它音段截取的信噪比较高的音调波形替代该音段的音调波形，这可以形成信噪比较高的波形数据库。

图 8 是一示意图，例示了在图 3 之步骤 330 中从一个音调波形组的音调波形中选择一代表性音调波形的第二种方法。在图 8 中，标号 710、 720、 730、 740 和 750 是用音素分类获得的音调波形分组。在该情况下，从各组中选择音调波形是这样做的，以致于所选的音调波形具有相似的相位特征。例如在图 8 中，从每个组中选择一个正峰值位于其中心的音调波形。也就是说，分别从组 710、 720、 730、 740 和 750 中选择音调波形 714、 722、 733、 743 和 751。应该注意，可以通过用例如傅里叶变换等手段分析每个音调波形的相位特征，作更精确的选择。

即使从不同的音段中收集音调波形，用这样方法选择代表性音调波形仍能使相位特征相似的音调波形合并起来，这可以避免因相位特征不同而引起的声音质量的劣化。

在以上描述中，每个音段只有一个值，因此每个音调波形没有音调变化。如

果只根据语音的文本数据合成语音，那么这可能是足够的。 但是，如果进行语音合成不仅要根据文本数据而且要根据语音的音调信息以便提供更自然的合成语音，那么以下所述的波形数据库将更好。

较佳的波形数据库

图 9 是一示意图，示出了依照本发明一较佳实施例的有声发音波数据库的结构．在图 9 中，有声发音波形数据库 900 包括一个音调波形指针表组 960 和多个用诸如功率频谱等音素分类而成的音调波形表组{365π|(π 表示语言中所用的音素，即π=a 、 i 、 u 、 e 、 o 、 k 、 s ，……)}．每个音调波形表组 365π 例如 365a 包括预定音调(频率)带----200-250 赫兹， 250-300 赫兹， 300-350 赫兹，… 的音调波形表 365a1 ， 365a2 ， 365a3 ， … 365aN ，其中 N 是预定音调带的数目。每个音调波形表 365πα(α=1 ， 2 ，… ， N)的结构与图 4B 中音调波形表 365 的结构相同。(‘α’ 是音调带编号。例如α=1 表示音调带 200-250 赫兹，α=2 表示音调带 250-300 赫兹，等等。)可以用任何形式实现用音素进行的分类或分组，例如通过实际将同一分组的音调波形表 365π1 至 365πN 存储在一个相关的折叠式表格或目录中，或者通过使用一个将音素 ‘π’ 和音调带 ‘α’ 的信息与相应的音调波形表 365πα 相联系的表。

图 10 例示了一个音调波形指针表的结构，例如图 9 所示的(关于音素链接格式 ‘ inu ’)960inu 。对于每个音素链格式，产生一个音调波形指针表。在图 10 中，除了已将记录 ID 从音素链接格式(音段)ID 改变至音调(频率)带，音素波形指针表 960inu 几乎与图 4A 的音调波形指针表 360 相同。诸如 ‘ i100 ’ 和 ‘ n100 ’ 等表达式表示音调波形 ID 。

在图 4A 和图 4B 的有声发音波形数据库中，每个音素链接格式只有一个音段。但是，在图 9 和图 10 的有声发音波长数据库 900 中，每个音素链接格式有四个音段。为此，以下必须区别音素链接格式和音段。将每个音素链接格式的 ID 表示成 IDp ， p=1 ， 2 ，… ， P ，其中 P 是(以下描述的)一个抽样组中音素链接格式的数目。以下运用变量 ‘ p ’ ，用 960p 表示一个音素链接格式 IDp 的音调波形指针表。

有一行(水平)数值，每个数值表示该数值所在的列中音调波形结束时的消逝时间。带阴影的音调波形 ID 是源于该音调波形指针表 960p 之音素链接格式(IDp)

中其个音段的音调波形的 ID，或是与那些音调波形接近相似并因此从其它音段上截取的音调波形的 ID。因此，一列中总有一个带阴影的音调波形 ID。但是，不能保证其余音调波形 ID 字段中也存在音调波形 ID，即在某些其余的音调波形 ID 字段中可以没有 ID。如果要参照空的音调波形 ID 字段，则最好参照一个相邻的 ID 字段。每个音调波形指针表 960p 仍有标志信息字段。图 10 所示的标志信息是最简单的例子，具有与图 4A 相同的结构。

图 11 是一流程图，示出了形成图 9 中有声发音波形数据库 900 的过程。在图 11 中，在步骤 800，如此准备一组音段抽样，使得每个预定音调带中包括每个音素链接格式 IDp。在步骤 810，将每个音段分成音调波形。在步骤 820，用音素将音调波形归类到音素组中，再将每个音素组归类到预定音调带的音调组中。在步骤 830，将每个音调组的音调波形归类到音调波形接近相似的分组中。在步骤 840，从每个组中选择一个音调波形，并为该选中的音调波形(或该分组)指定一个 ID。在步骤 850，建立每个音调带所选波形组的一个音调波形表。然后在步骤 860，对每个音素链接格式，产生一个音调波形指针表，在该表中每条记录至少包括音调带数据以及构成由音调带数据确定的音调带之音段(格式)的音调波形的 ID。

哑声波形表

对于包括一个哑声(辅音)的每个音素链接(例如，VCV 链接)音段，如果将哑声波形存储在一个波形表中，那么这会使该表（或数据库）冗长。用有声发音情况下所用的相同方式可以避免该现象。

图 12 是一示意图，示出了不同音段是如何共享一个通用哑声的。在图 12 中，与只包含有声发音的音段的情况相同，将音段 ‘ aka ’ 1102 分成音调波形 1110，…，1112，哑声 1115 和音调波形 1118，…，1119，而将音段 ‘ ika ’ 1105 分成音调波形 1120，…，1122，哑声 1125 和音调波形 1128，…，1129。在该情况下，两个音段 ‘aka’ 1102 和 ‘ika’ 1105 共享无声辅音 1115 和 1125。

图 13 是一流程图，示出了依照本发明说明性实施例形成一哑声波形表的过程。在图 13 中，在步骤 1300，准备一组包含一个哑声的音段抽样。在步骤 1310，从各音段中收集哑声。在步骤 1320，将哑声归类到接近相似的哑声组中。在步骤 1330，从每个组中选择一个哑声(波形)，并为该选中的哑声(或该组)指定一个

ID。在步骤 1340，产生一个哑声波形表，表中的每条记录都包括一个指令的 ID 和由该 ID 识别的被选中的哑声波形。

语音合成系统的工作情况

图 14 是一流程图，例示了图 4 中使用有声发音波形数据库的语音合成程序的流程。进入该程序时，在步骤 1400，控制器 10 接收要合成的语音的文本数据。在步骤 1410，控制器 10 确定合成该语音所必须的诸音段的音素链接格式；并计算包括持续时间和功率型谱的节律(或韵律)。在步骤 1420，控制器 10 从图 4A 的音调波形指针表 360 中获取每个被确定音素链接格式所用的音调波形 ID。在步骤 1430，控制器 10 从音调波形表 365 中获取与所得 ID 相关的音调波形，并从常规的哑声波形表中获取哑声波形，然后用所得的波形合成各音段。接着在步骤 1440，控制器 10 合并合成后的音段，产生合成语音，并结束程序。

图 15 是一流程图，例示了图 9 和图 10 中使用有声发音波形数据库的语音合成程序的流程。图 15 的步骤 1400 和步骤 1440 与图 14 中的相同。因此，此处只描述步骤 1510 至 1530。在步骤 1510，响应于接收到文本数据或语音符号数据(phonetic sign data)，控制器 10 确定合成该语音所必须的每个音段的音素链接格式(IDp)和音调带(α)，并计算包括该语音之持续时间和功率型谱的节律(或韵律)信息。在步骤 1520，控制器 10 根据计算得到的韵律信息，从图 10 所示的音调波形指针表 960IDp 中获取被确定音调带(α)中每个音段所用的音调波形 ID。在步骤 1530，控制器 10 从音调波形表 365πα 中获取与所得 ID 相关的音调波形，并从常规的哑声波形表中获取哑声波形，然后用所得的波形合成各音段。接着在步骤 1440，控制器 10 合并合成后的音段，产生合成语音，并结束程序。

不脱离本发明的精神和范围可以构造许多本发明的不同实施例。应该理解，除非如后附的权利要求限定的，本发明不局限于说明书中描述的具体实施例。

说 明 书 附 图

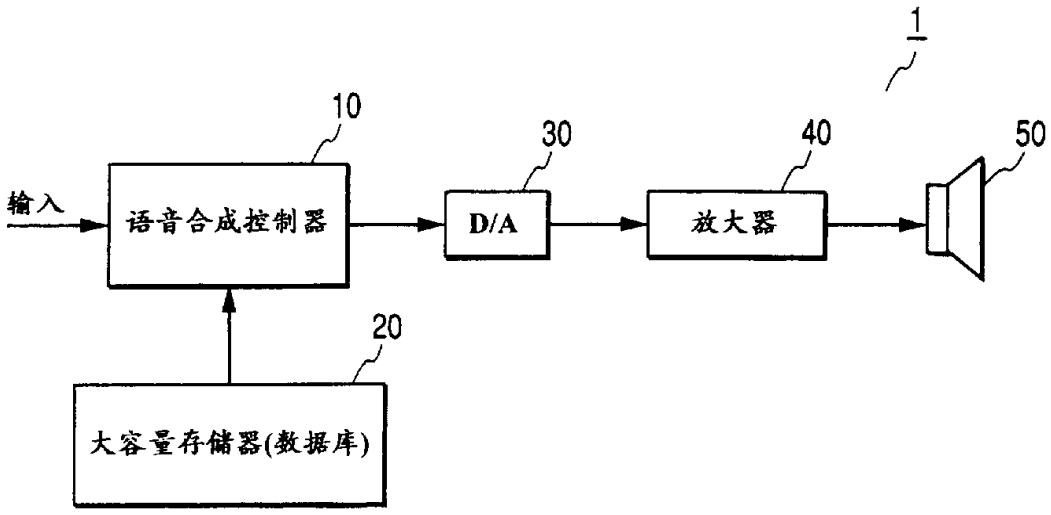


图 1

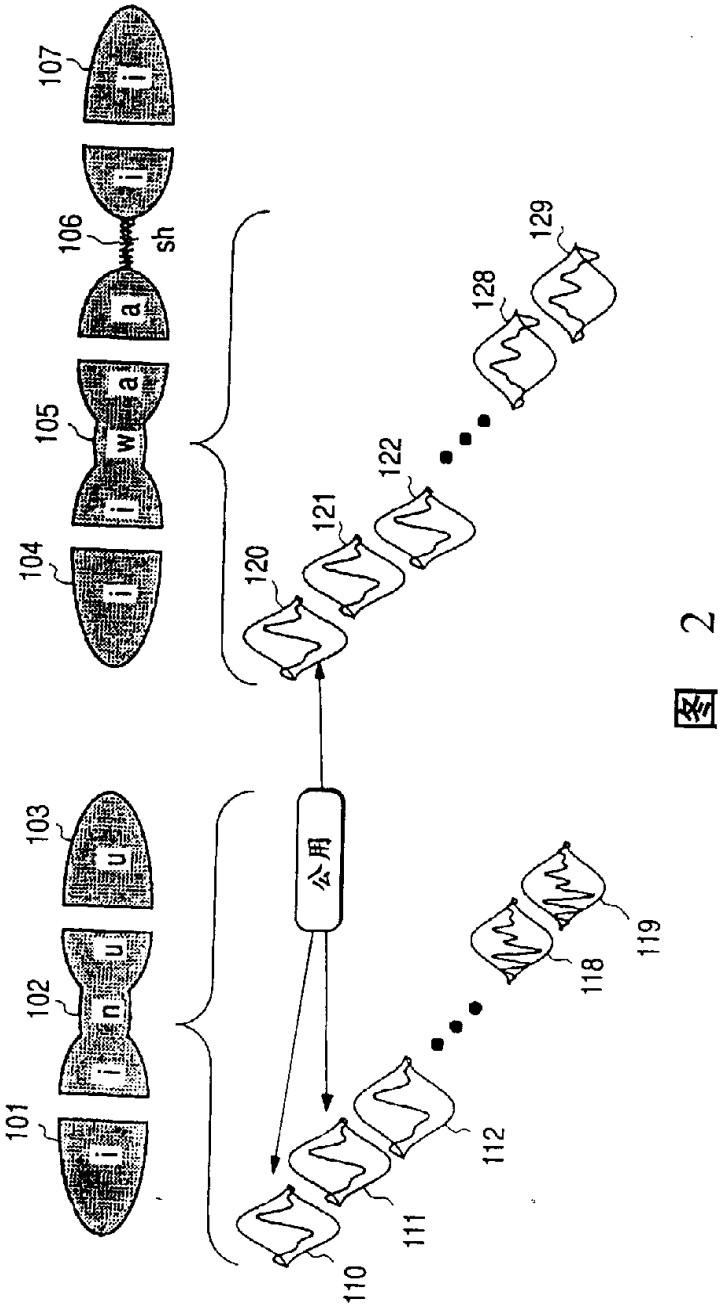

101
i
102
i
n
u
103
u
公用
104
i
105
i
w
a
106
a
sh
i
107
i
110
111
112
118
119
120
121
122
128
129

图 2

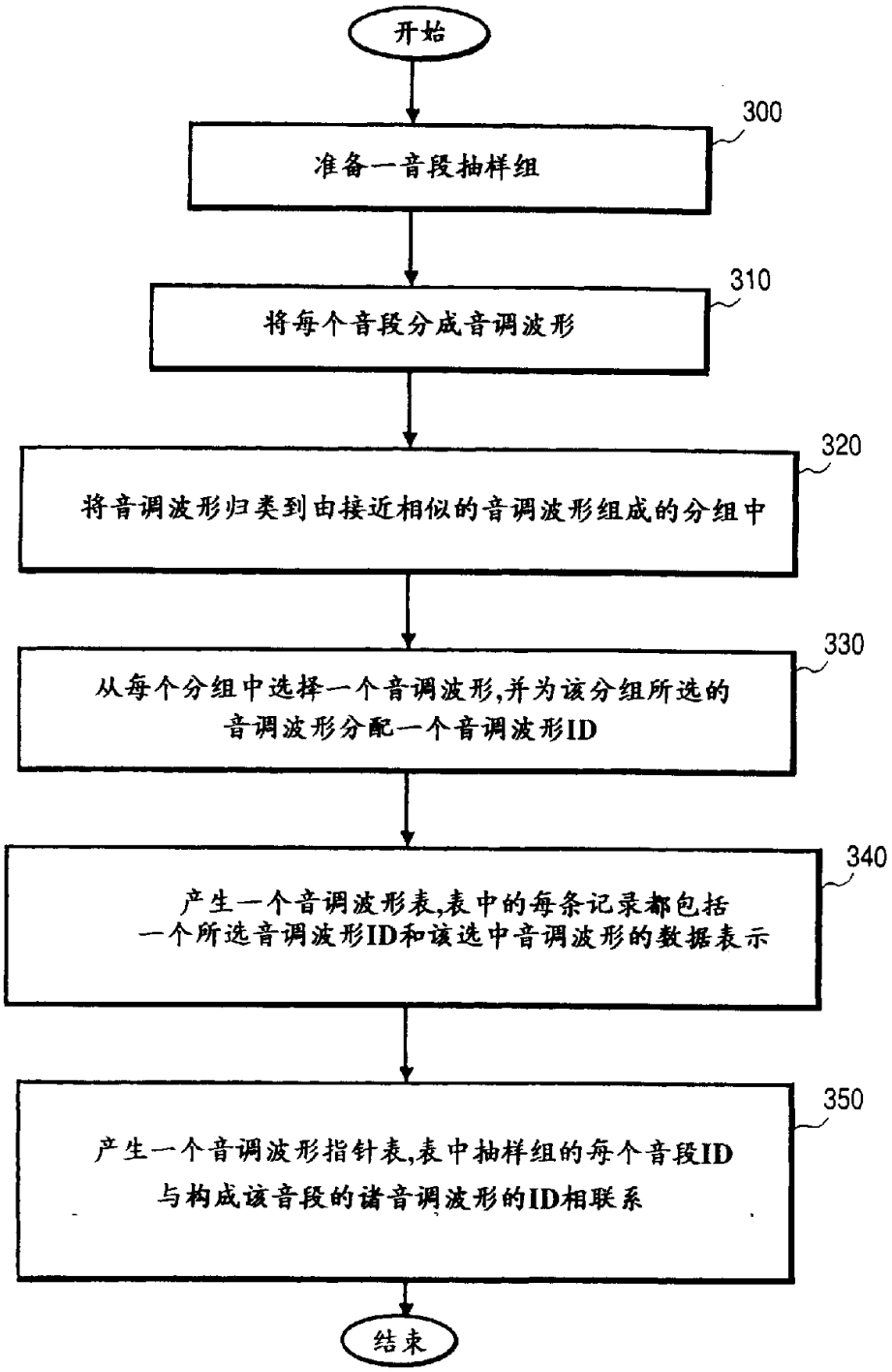

开始
300
准备一音段抽样组
310
将每个音段分成音调波形
320
将音调波形归类到由接近相似的音调波形组成的分组中
330
从每个分组中选择一个音调波形,并为该分组所选的
音调波形分配一个音调波形ID
340
产生一个音调波形表,表中的每条记录都包括
一个所选音调波形ID和该选中音调波形的数据表示
350
产生一个音调波形指针表,表中抽样组的每个音段ID
与构成该音段的诸音调波形的ID相联系
结束

图 3

360

音段ID	音调波形ID											标志信息		
												前元音的音调计数	辅音的音调计数	后元音的音调计数
:														
inu	i100	i100	i101	…	n100	n101	…	u200	u200	u201	…	20	10	20
:														
iwa	i100	i200	i300	…	w30	w10	…	a500	a400	a300	…	23	15	30
:														

图 4A

365

音调波形ID	音调波形数据

图 4B

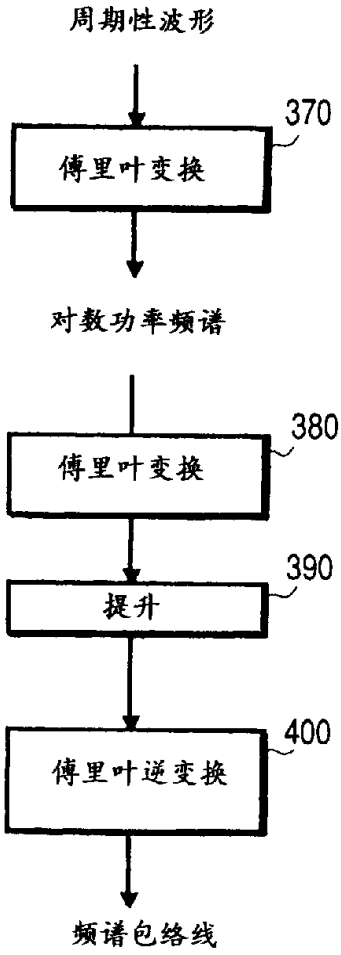

周期性波形
370
傅里叶变换
对数功率频谱
380
傅里叶变换
390
提升
400
傅里叶逆变换
频谱包络线

图 5A

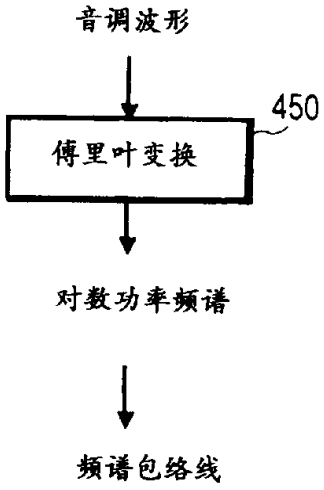

音调波形
450
傅里叶变换
对数功率频谱
频谱包络线

图 5B

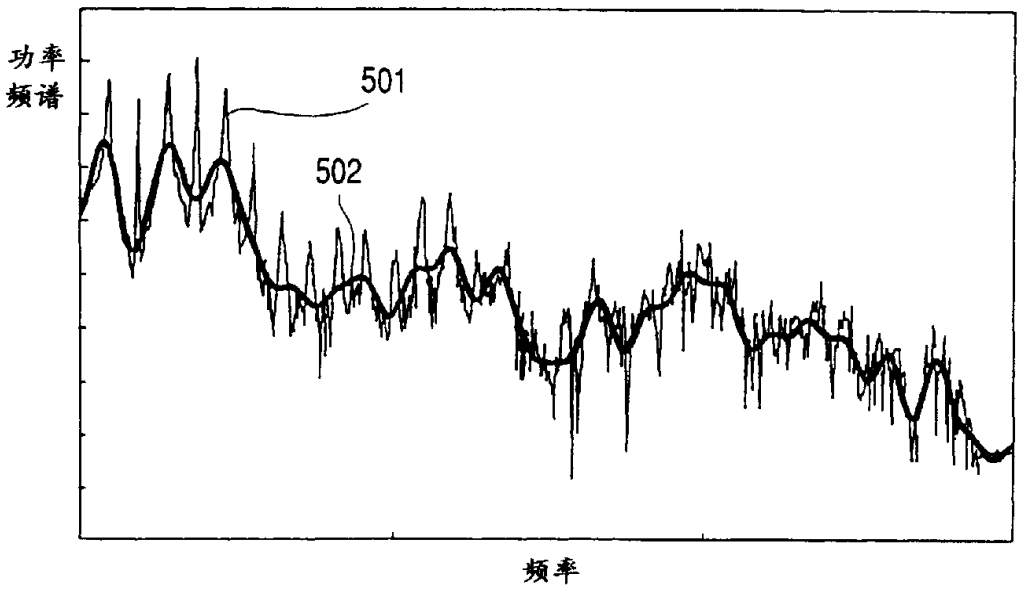


图 6

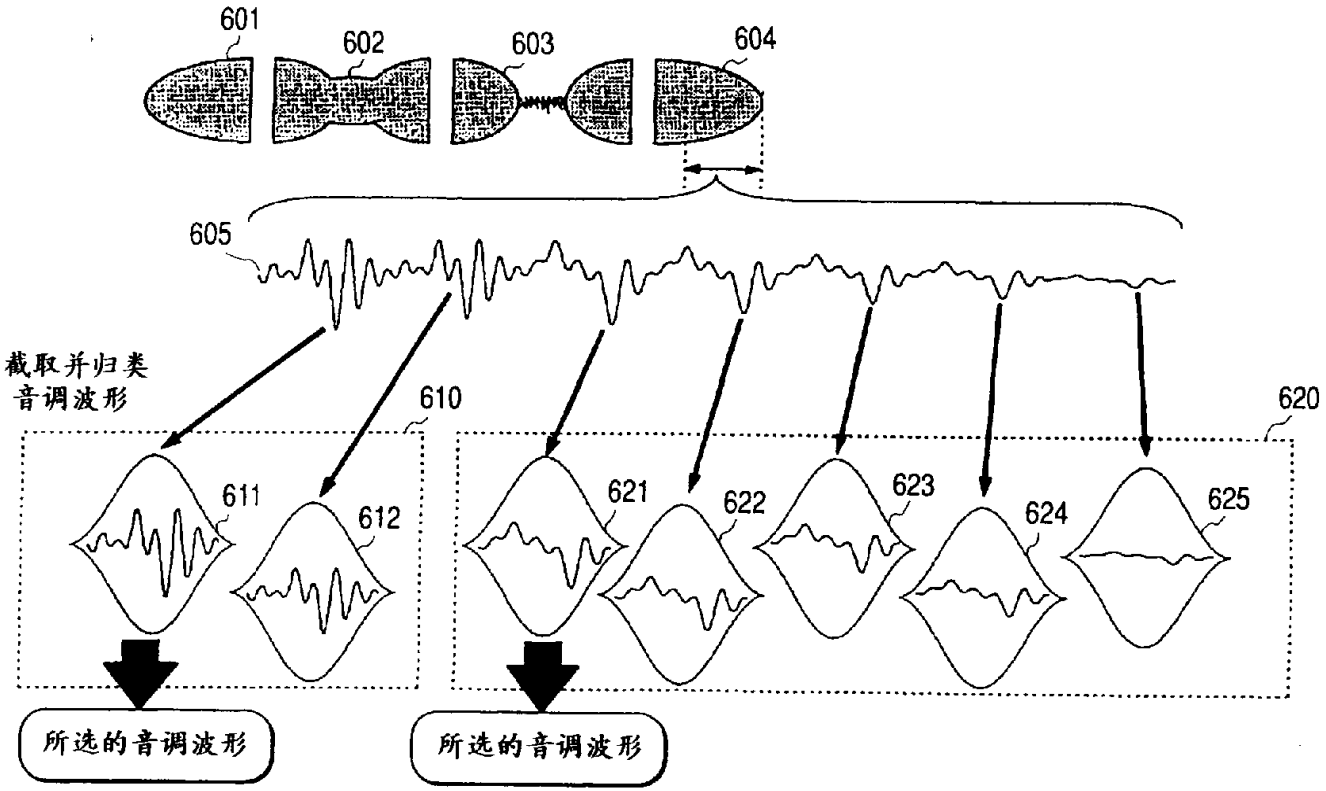

601
602
603
604
605
截取并归类
音调波形
610
620
611
612
621
622
623
624
625
所选的音调波形
所选的音调波形

图 7

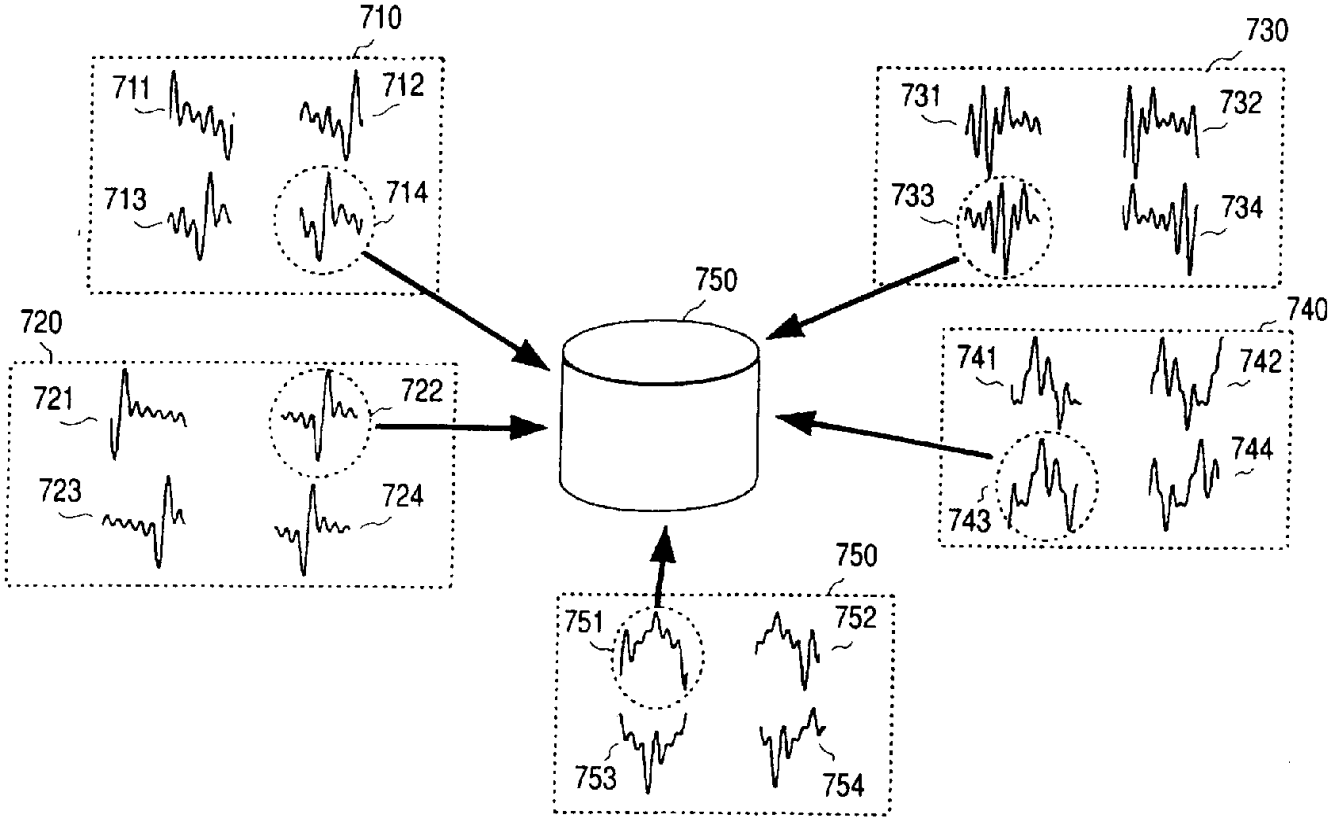

710
711
712
713
714
730
731
732
733
734
750
720
721
722
723
724
740
741
742
743
744
750
751
752
753
754

图 8

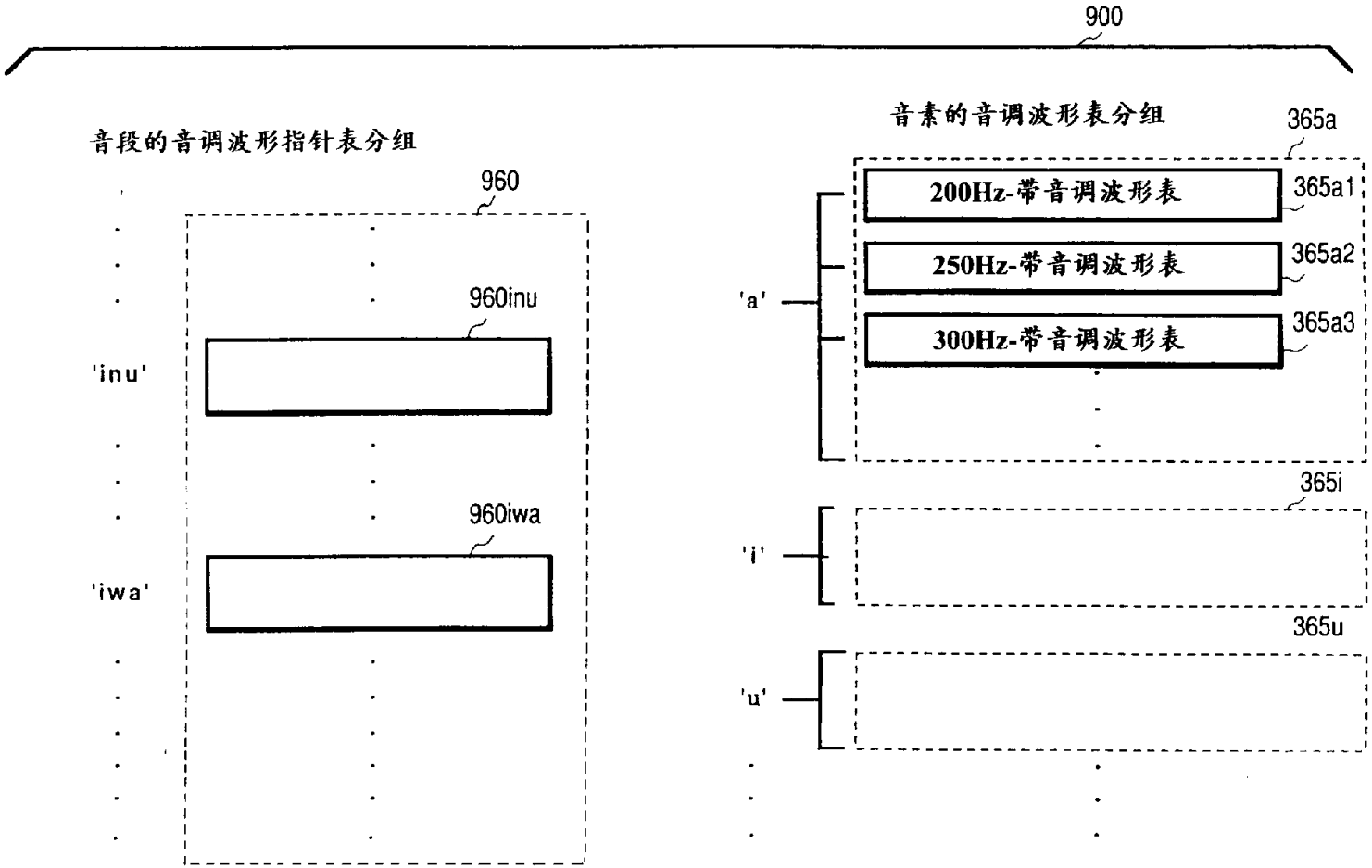

900
音段的音调波形指针表分组
960
960inu
'inu'
960iwa
'iwa'
音素的音调波形表分组
365a
200Hz-带音调波形表
365a1
250Hz-带音调波形表
365a2
300Hz-带音调波形表
365a3
'a'
365i
'i'
365u
'u'

图 9

960inu

消逝时间(毫秒) / 音调带(赫兹)	2.4	4.5	6.4	…	25.4	27.1	…	32.2	34.1	36.3	38.9	…	音调计数		
													前元音	辅音	后元音
350～400	i130	i120	i121	…	n130	n121	…	u230	u220	u221	u221	…	20	10	20
300～350	i120		i121	…		n121	…	u220	u220	u221	u221	…			
250～300	i110	i110	i111	…	n110	n101	…	u210		u221	u221	…			
250～250	i100	i100	i101	…	n100	n101	…	u200	u200	u221	u221	…			

图 10

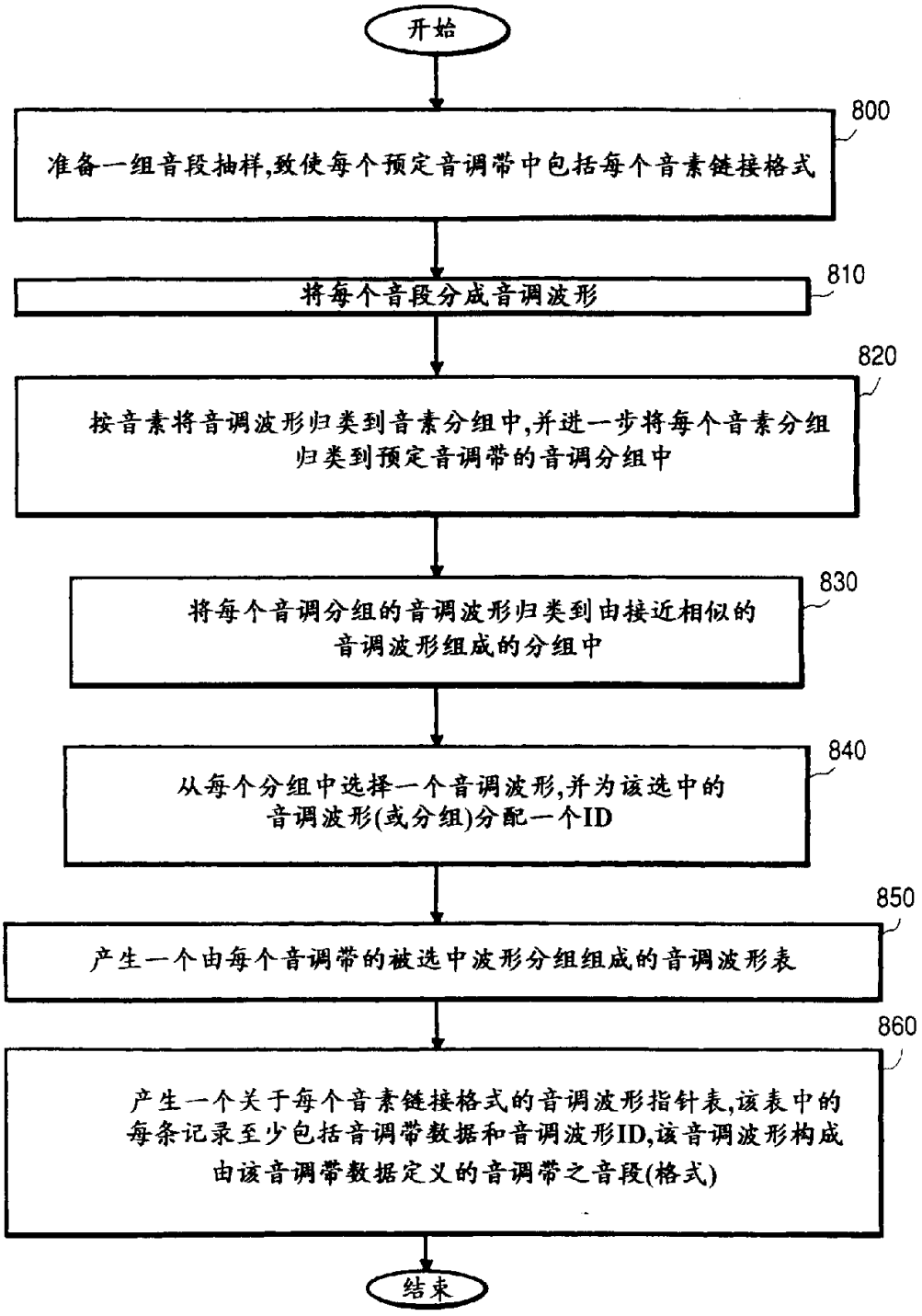

开始
800
准备一组音段抽样,致使每个预定音调带中包括每个音素链接格式
810
将每个音段分成音调波形
820
按音素将音调波形归类到音素分组中,并进一步将每个音素分组归类到预定音调带的音调分组中
830
将每个音调分组的音调波形归类到由接近相似的音调波形组成的分组中
840
从每个分组中选择一个音调波形,并为该选中的音调波形(或分组)分配一个ID
850
产生一个由每个音调带的被选中波形分组组成的音调波形表
860
产生一个关于每个音素链接格式的音调波形指针表,该表中的每条记录至少包括音调带数据和音调波形ID,该音调波形构成由该音调带数据定义的音调带之音段(格式)
结束

图 11

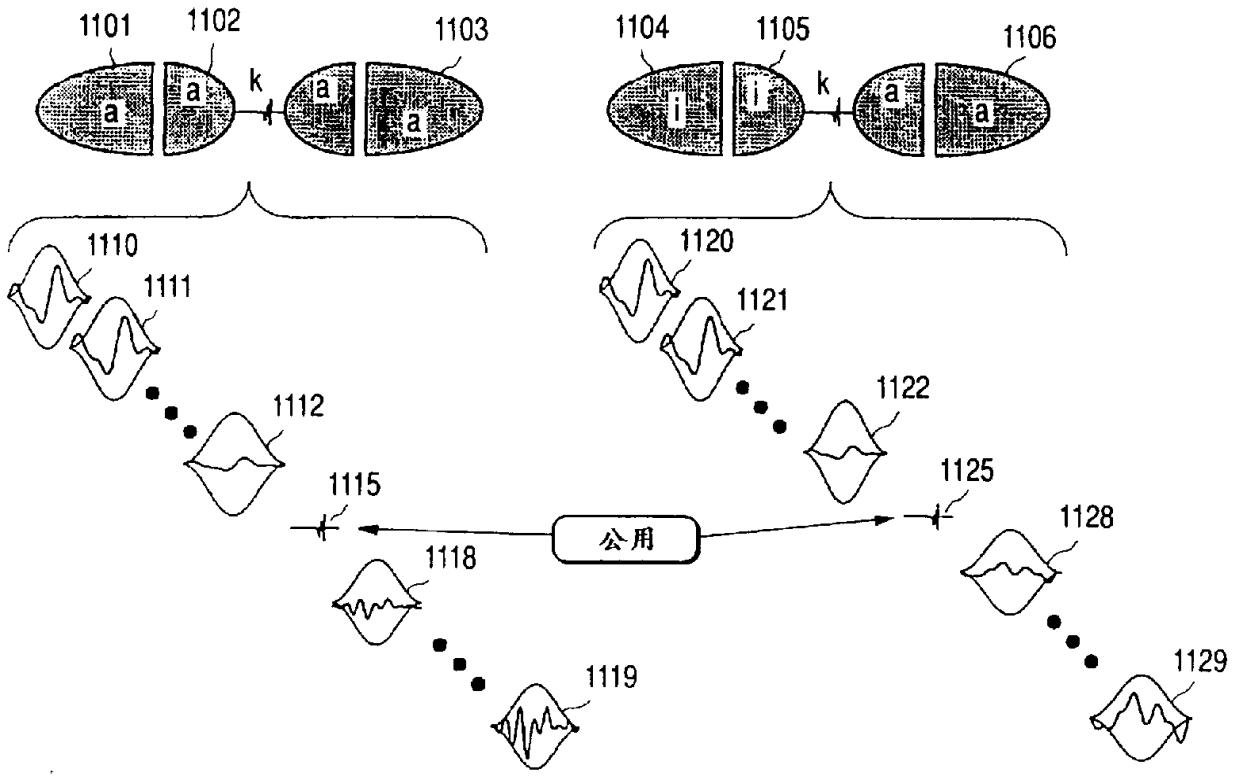

1101
1102
1103
1104
1105
1106
a
a
k
a
a
i
i
k
a
a
1110
1111
1112
1120
1121
1122
1115
1125
公用
1118
1119
1128
1129

图 12

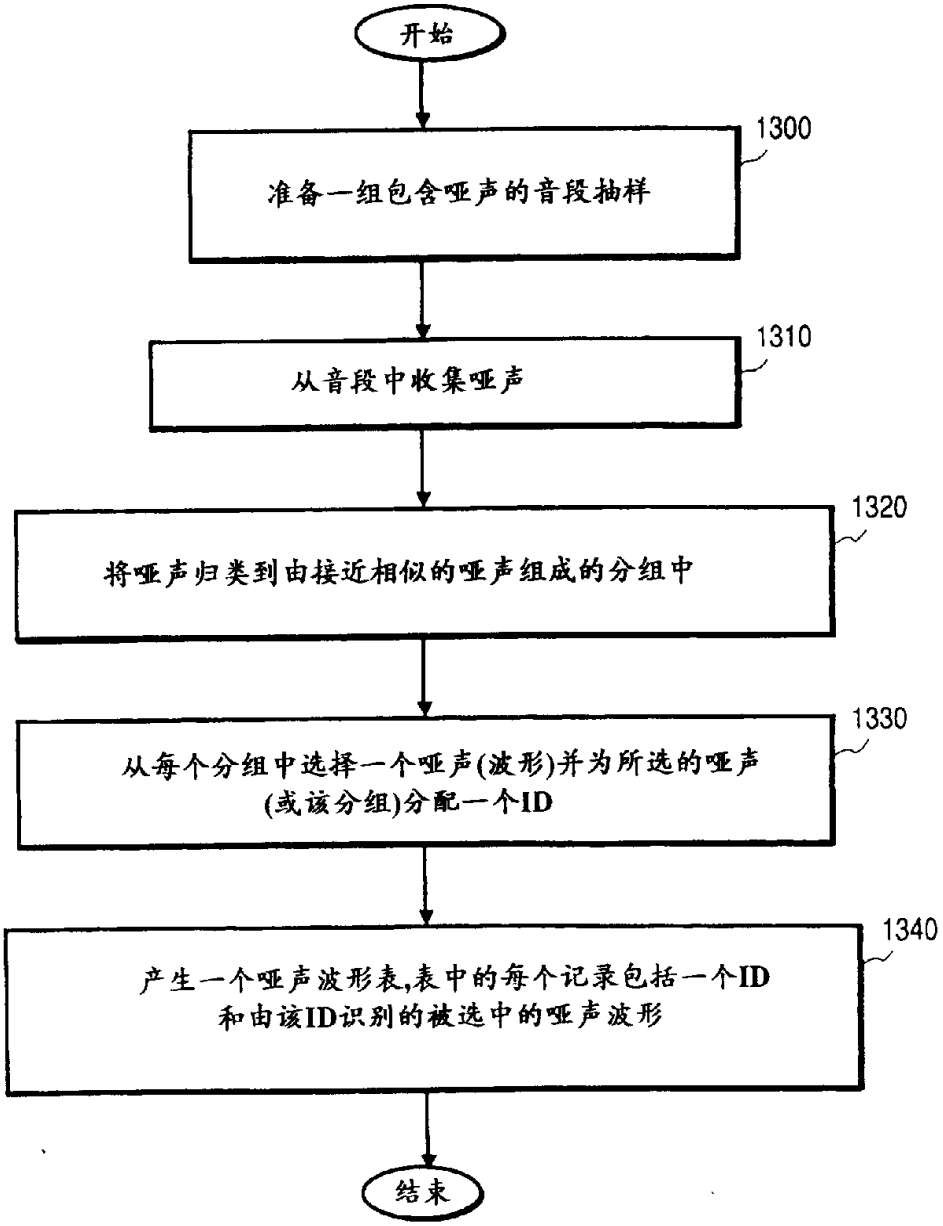

开始
1300
准备一组包含哑声的音段抽样
1310
从音段中收集哑声
1320
将哑声归类到由接近相似的哑声组成的分组中
1330
从每个分组中选择一个哑声(波形)并为所选的哑声(或该分组)分配一个ID
1340
产生一个哑声波形表,表中的每个记录包括一个ID和由该ID识别的被选中的哑声波形
结束

图 13

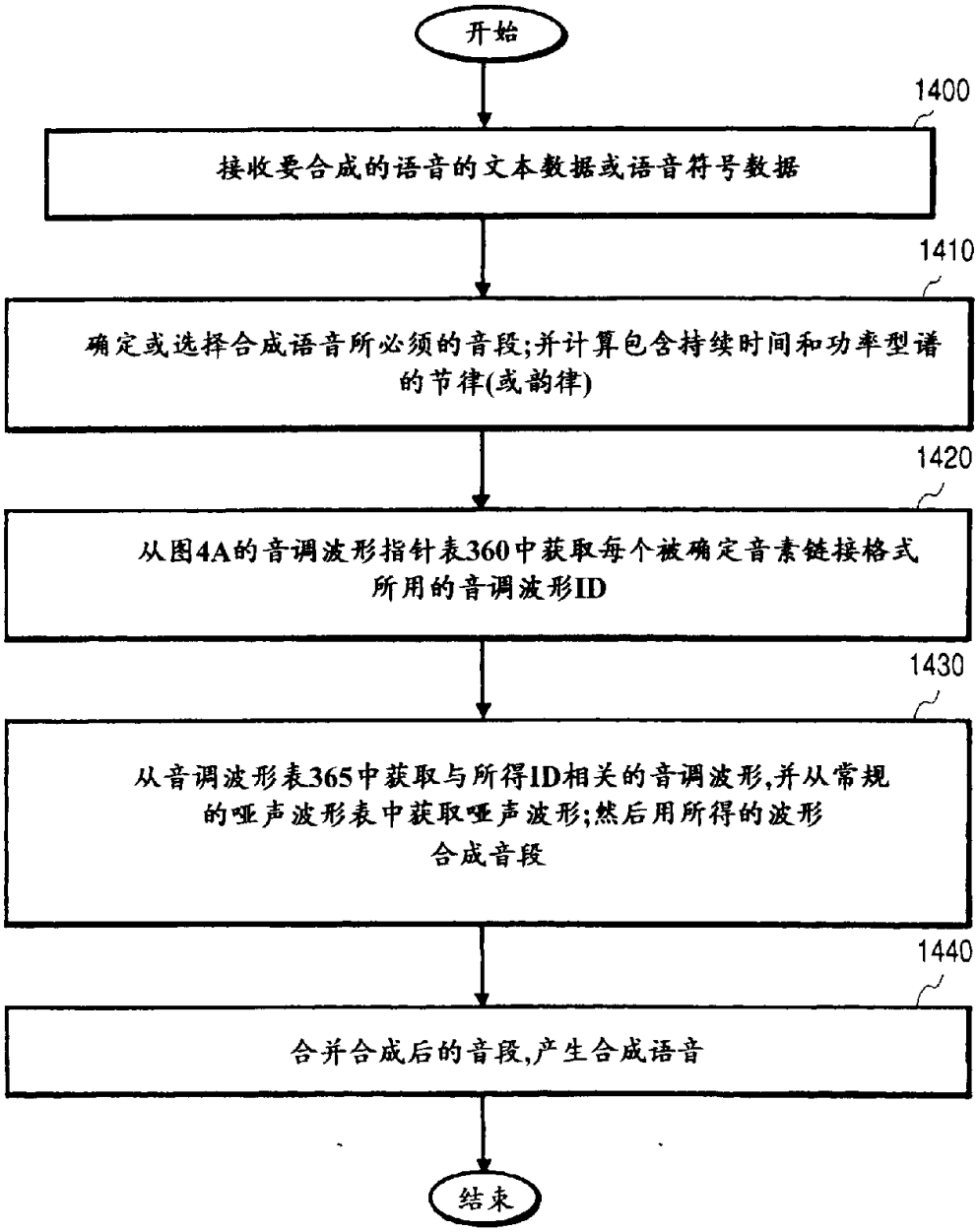

开始
1400
接收要合成的语音的文本数据或语音符号数据
1410
确定或选择合成语音所必须的音段;并计算包含持续时间和功率型谱的节律(或韵律)
1420
从图4A的音调波形指针表360中获取每个被确定音素链接格式所用的音调波形ID
1430
从音调波形表365中获取与所得ID相关的音调波形,并从常规的哑声波形表中获取哑声波形;然后用所得的波形合成音段
1440
合并合成后的音段,产生合成语音
结束

图 14

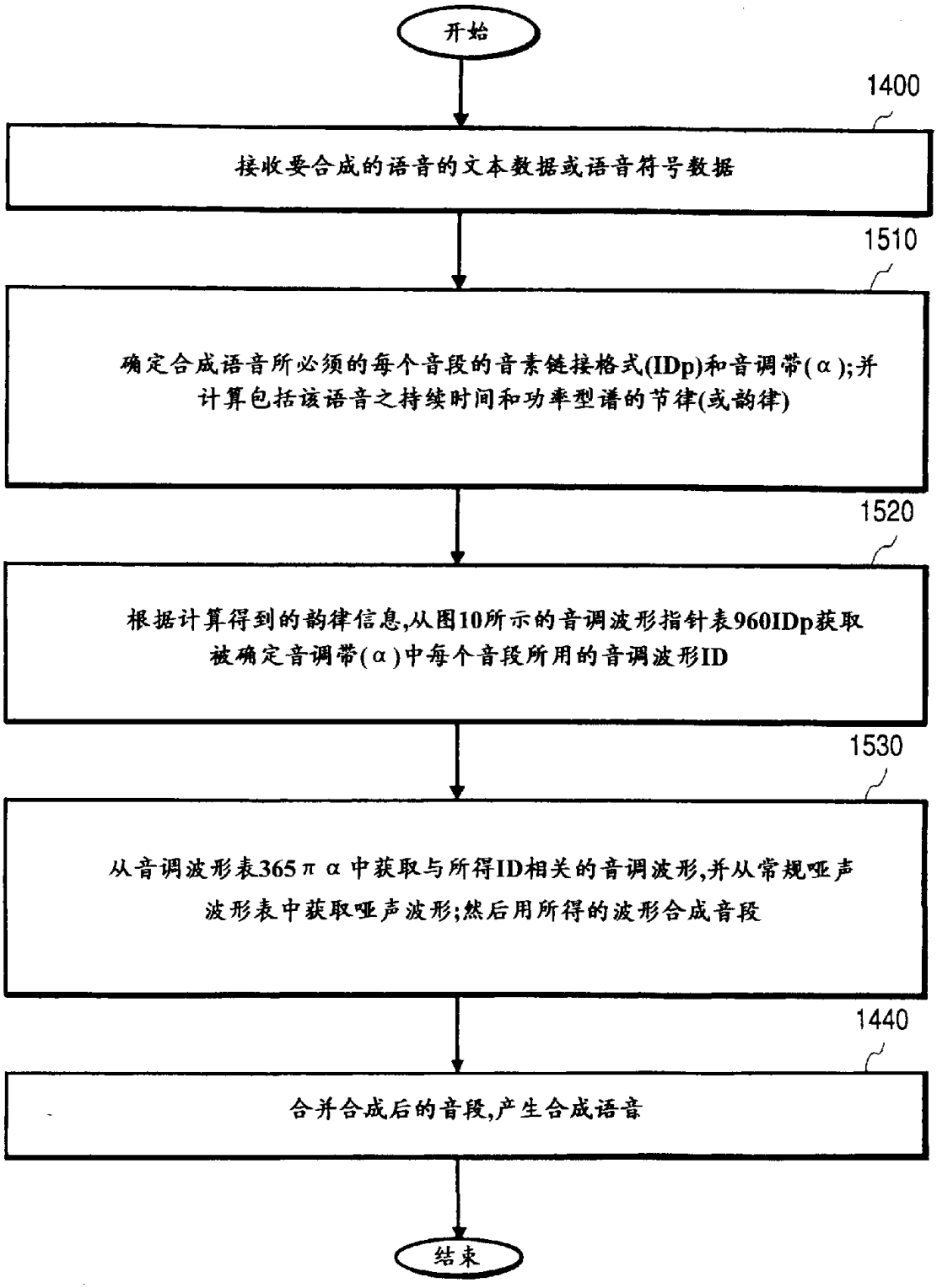

开始
1400
接收要合成的语音的文本数据或语音符号数据
1510
确定合成语音所必须的每个音段的音素链接格式(IDp)和音调带(α);并
计算包括该语音之持续时间和功率型谱的节律(或韵律)
1520
根据计算得到的韵律信息,从图10所示的音调波形指针表960IDp获取
被确定音调带(α)中每个音段所用的音调波形ID
1530
从音调波形表365πα中获取与所得ID相关的音调波形,并从常规哑声
波形表中获取哑声波形;然后用所得的波形合成音段
1440
合并合成后的音段,产生合成语音
结束

图 15