

[19] 中华人民共和国国家知识产权局

[51] Int. Cl.

G10L 13/00 (2006.01)

H04M 11/00 (2006.01)



[12] 发明专利说明书

专利号 ZL 200410047414.6

[45] 授权公告日 2007 年 4 月 11 日

[11] 授权公告号 CN 1310209C

[22] 申请日 2004.5.28

[21] 申请号 200410047414.6

[30] 优先权

[32] 2003. 5. 29 [33] JP [31] 2003 - 152895

[32] 2003. 9. 30 [33] JP [31] 2003 - 340171

[73] 专利权人 雅马哈株式会社

地址 日本静岡県

[72] 发明人 川岛隆宏

[56] 参考文献

JP2002221980A 2002.8.9

JP2002366186A 2002.12.20

CN1416053A 2003.5.7

JP200329774A 2003.1.31

JP200151688A 2001.2.23

审查员 刘亚斌

[74] 专利代理机构 隆天国际知识产权代理有限公司

代理人 经志强 潘培坤

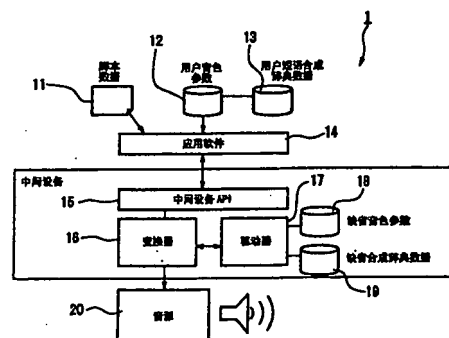
权利要求书 3 页 说明书 21 页 附图 15 页

[54] 发明名称

语音和乐曲再生装置

[57] 摘要

本发明所涉及的语音・乐曲再生装置由再生合成语音信号等的中间设备、根据该合成语音信号再生所希望的语音或乐曲的音源(20)和扬声器构成;其中中间设备把脚本数据(11)、用户音色参数(12)及用户短语合成辞典数据(13)联合起来的同时,根据缺省音色参数(18)和缺省合成辞典数据(19)再生合成语音信号等。在使用记述种种事件的 HV - 脚本作为脚本数据的情况下,根据其事件的类别适宜地把所希望的波形数据、包含音符信息的乐曲短句数据以及共振峰帧数据组合起来再生。



1. 一种语音再生装置，其具有存储装置、登录装置、输入装置和语音合成装置；其中，

该存储装置存储合成辞典数据，该合成辞典数据将对应于表示规定发音单元的发音文字的共振峰帧数据与该发音文字对应起来并预先进行保存；

该登录装置按照用户的操作将用户短语数据登录到用户辞典数据中，该用户短语数据表示用于代替与保存在前述合成辞典数据中的发音文字相对应的共振峰帧数据而使用的其他共振峰帧数据；

该输入装置输入包含由多个发音文字构成的字符串和事件数据的脚本数据，该事件数据指示对应于该字符串的至少一部分发音文字的共振峰帧数据的置换；

该语音合成装置解释输入的所述脚本数据，并根据上述字符串中的至少一部分以外的发音文字从所述合成辞典数据中读出共振峰帧数据，根据所述事件数据以及所述一部分字符串从所述用户辞典数据读出所述用户短语数据，再根据所读出的所述共振峰帧数据以及所读出的所述用户短语数据生成合成语音。

2. 根据权利要求1的语音再生装置，其特征在于，进一步具有根据乐曲再生信息再生乐曲的乐曲再生装置；

在所述输入装置中输入数据交换格式，所述数据交换格式是包含用于再生乐曲的乐曲再生信息和含有所述脚本数据的音乐再生信息和所述用户短语数据、且同步再生基于所述乐曲再生信息的乐曲再生和基于所述语音再生信息的语音再生的信息结构；

所述乐曲再生装置再生包含在所述数据交换格式中的所述乐曲再生信息；

所述语音合成装置再生包含在所述数据交换格式中的所述语音再生信息。

3. 一种便携式终端装置，具备权利要求1或2所述的语音再生装置。

4. 一种语音再生装置，由存储声音数据的第一存储装置、存储脚本数据的第二存储装置、再生指示装置、合成发音信号生成装置、声音信号生成装置和合成语音生成装置构成；其中，

该脚本数据记述了由表示规定发音单元的发音文字构成的字符串和指示所述声音数据的再生的事件数据；

该再生指示装置从所述第二存储装置中读出所述脚本数据, 根据所述脚本数据中的字符串指示该发音, 根据所述脚本数据中的事件数据指示所述声音数据的再生;

该合成发音信号生成装置根据来自所述再生指示装置的字符串的发音指示, 进行语音合成, 并生成合成发音信号;

该声音信号生成装置根据来自所述再生指示装置的所述声音数据的再生指示从所述第一存储装置中读出所述声音数据, 根据该声音数据生成声音信号;

该合成语音装置根据所述合成发音信号生成合成语音, 根据所述声音信号生成声音。

5. 根据权利要求4的语音再生装置, 其特征在于, 所述声音数据是通过采样规定的声音生成的波形数据。

6. 根据权利要求4的语音再生装置, 其特征在于, 所述声音数据是包含表示应发音的声音的音调和音量的音符信息的乐曲数据。

7. 根据权利要求4的语音再生装置, 其特征在于, 所述合成发音信号生成装置存储有使文字的发音带有特征的共振峰控制参数, 用对应于所述脚本数据中的字符串的共振峰控制参数进行语音合成。

8. 根据权利要求4至7中任意一项所述的语音再生装置, 其特征在于, 所述脚本数据用由文本数据构成的文件记述。

9. 根据权利要求4的语音再生装置, 其特征在于, 设置将对应于表示规定发音单元的发音文字的共振峰帧数据与该发音文字对应起来并预先保存的合成辞典数据;

根据用户的操作, 将用户短语数据登录到用户辞典数据中, 该用户短语数据表示用于代替与保存在前述合成辞典数据中的发音文字相对应的共振峰帧数据而使用的其他共振峰帧数据;

在所述脚本数据包含指示对应于所述字符串的至少一部分发音文字的共振峰帧数据的置换的事件数据的情况下, 所述合成发音信号生成装置根据所述一部分发音文字以外的发音文字从所述合成辞典数据中读出共振峰帧数据, 并根据所述事件数据以及所述一部分字符串从所述用户辞典数据读出所述用户短语数据, 再根据所读出的所述共振峰帧数据以及所读出的所述用户

短语数据生成所述合成发音信号。

10. 一种便携式终端装置，具备权利要求 4 至 9 中任意一项记载的语音再生装置。

语音和乐曲再生装置

技术领域

本发明涉及语音和乐曲再生装置,特别是涉及通过语音合成再生特定的话的同时,把字符信息变换为语音或乐曲再生出来的语音和乐曲再生装置。

背景技术

以往,设计有把电子邮件等字符串(character string information)信息变换为语音输出的字符串语音变换装置。日本公开专利特开 2001-7937 号展示出现有的字符串语音变换装置之一例,其中以文节为单元来划分字符串信息,输出语音的同时,将其内容显示在显示器上。

公知的方法有再生对乐曲短句或语音短语等进行采样作成的波形数据(或采样数据)的方法,或由 SMF(标准 MIDI 文件)或 SMAF(合成音乐移动应用文件:synthetic music mobile application file)等音符信息构成一个乐曲短句再再生该乐曲短句的方法。例如,日本公开专利特开 2001-51688 号中披露了一种电子邮件读取装置,能够把电子邮件中的字符串信息与乐音信息分离开,再各自发音再生出来。

但是,在现有的字符串语音变换装置中,由于是以文节(句子(clause)或短语)为单元来划分字符串信息,进行语音输出,所以该语音输出是发音单元(或字符单元)的语音的集合,在再生该发音单元的连接点时,与正常的说话语音相比,听取者就感觉不和谐。即,在现有的字符串语音变换装置中,就文节整体而言,不能以音质良好的语音变化并输出该音色,换言之,不能够输出接近于人说话语言的自然的语音。

为解决上述问题所考虑的方法是例如预先对每个文节(以下称为“短语”)的语音采样,并作为语音数据保存起来,再生时作为相应的语音波形进行输出。但是,为了提高语音输出的品质,这种方法必须提高采样频率,因此,就必须保存大容量的语音数据,在便携式电话机(手提电话或移动电话)等存储容量比较有限的装置中就会伴随出现技术性的困难。

另外，再生对乐曲或语音采样作成的波形数据的所述现有的方法或由 SMF 或 SMAF 等音符信息构成一个乐曲数据再再生该乐曲数据的现有方法中，乐曲或语音的再生定时并不是按文本文件形式记述的，因此，很难按照用户的意图进行基于字符串信息的语音再生与波形数据再生或乐曲数据再生之间的组合。

发明内容

为解决上述的问题，本发明的目的在于提供一种语音再生装置，能够把由字符串信息等构成的所希望的文节（或短语）作成音质良好的语音，并使其音色变化，再生输出。

本发明的其他目的在于提供一种语音・乐曲再生装置，用户能够简单地进行语音再生或波形数据再生与乐曲数据再生的组合，因此，能够进行忠实于用户的意图的语音和乐曲的再生。

按照本发明的语音再生装置把由对应于预先规定的发音单元的共振峰帧数据构成的数据库作为合成辞典数据存储起来，当赋予有关并列构成发音单元的字符串的信息时，用所述合成辞典数据进行语音合成。这里，把共振峰帧数据置换为任意的用户短语数据，赋予字符串信息时，用置换为该用户短语数据的合成辞典数据进行语音合成。把加工共振峰帧数据的音色参数附加在用户短语数据中。另外，在语音合成时，使用包含用户短语数据的规定的交换格式。该数据交换格式是例如 SMAF 文件格式，不仅包含用户短语数据还可以包含各种块或乐曲再生信息。

具体地说，上述语音再生装置由保存对应于预先规定的发音单元的共振峰帧数据的缺省合成辞典数据、把该共振峰帧数据置换为用户短语数据的中间设备・应用程序・接口（API）、变换器、驱动器以及音源构成。这样，就能够把由字符串信息构成的所希望的短语作成音质良好的语音再生出来，而且能够使音色适宜地变化，进行再生。

涉及本发明的语音・乐曲再生装置存储有记述了文字的发音或预先存储的发音用数据的再生指示的脚本数据（即 HV-脚本）。根据该脚本数据生成对应于所述文字的语音信号并产生所希望的语音，同时生成对应于发音用数据的发音信号并产生所希望的语音或乐音。这里，发音用数据是例如由采样语音或乐曲所生成的波形数据构成，根据该波形数据生成合成发音信号。另外，在把发

音用数据作为包含音符信息的乐曲数据的情况下,根据该乐曲数据生成对应于音符信息的乐音信号。另外,在存储了使所述文字的发音具有特征的共振峰控制参数(或共振峰帧数据)的情况下,根据该共振峰控制参数生成语音信号。所述脚本数据也可以由用户任意作成,这时,脚本数据采取通过文本输入作成的规定的文件形式。

具体地说,在解释记述在 HV-脚本中的各种事件,并且该事件的类别表示波形数据的情况下,读出该波形数据并再生出来,另一方面,在事件的类别表示乐曲短句数据的情况下,进行该乐曲短句数据的再生处理。这时,根据乐曲短句数据中的时间信息读出并再生其音符数据。另外,与其他事件相对应,把用合成辞典数据输入的字符串变换为共振峰帧串,进行语音合成。这样,用户就能够把语音再生、波形数据再生和乐曲数据再生容易地组合起来进行。

附图说明

图 1 是表示本发明的第一实施例的语音再生装置的构成的框图。

图 2 是发音单元与短语 ID 的分配关系图。

图 3 是短语合成辞典数据的内容例。

图 4 是 SMAF 文件的格式例。

图 5 是表示 HV 创作工具的一例的功能框图。

图 6 是表示应用了语音再生装置的便携式通信终端的构成的框图。

图 7 是表示用户短语合成辞典数据的作成处理的流程图。

图 8 是表示用户短语合成辞典数据的再生处理的流程图。

图 9 是表示 SMAF 文件的作成处理的流程图。

图 10 是表示 SMAF 文件的再生处理的流程图。

图 11 是表示本发明的第二实施例的语音·乐曲再生装置的构成框图。

图 12 是各事件与波形数据以及乐曲短句数据之间的分配关系示例图。

图 13 是表示第二实施例的语音·乐曲再生处理的流程图。

图 14 是表示具备第二实施例的语音·乐曲再生装置的便携式电话的构成的框图。

图 15 是表示本发明的第三实施例的语音·乐曲再生装置的构成的框图。

图 16 是表示图 15 所示的语音·乐曲再生装置的动作的流程图。

具体实施方式

参照附图详细说明本发明的实施例。

图1是表示本发明的第一实施例的语音再生装置的构成的框图。

即,图1所示的语音再生装置1具备应用软件(application software)14、中间设备API(中间设备应用程序接口; middleware application program interface)15、变换器16、驱动器17、缺省音色参数(default tone color parameter)18、缺省合成辞典数据(default composition dictionary data)19和音源20,输入脚本数据(script data)11、用户音色参数12和用户短语合成辞典数据(user phrase composition dictionary data)13(可变长度)来再生语音。

语音再生装置1基本上采用根据按照用FM(frequency modulation)音源资源的CSM(composition sinusoidal modeling: 复合正弦波模型)语音合成方式的共振峰合成(formant composition)再生语音的方法。在本实施例中,定义用户短语合成辞典数据13,语音再生装置1参照该用户短语合成辞典数据13对音色参数以音素(phoneme)为单元分配用户短语。在音色参数中这样地分配有用用户短语合成辞典数据13的情况下,再生时,语音再生装置1把登录在缺省合成辞典数据内的音素置换为用户短语,然后根据该置换数据进行语音合成。而且,所谓上述的“音素”是发音的最小单元,在日本语的情况下,由元音(vowel)和辅音(consonant)构成。

下面说明语音再生装置1的详细构成。

图1中,脚本数据11是定义用来再生“HV(human voice;人类声音):用上述方法合成的语音”的数据格式的数据。即,脚本数据11表示包含含有韵律符号(intonation symbol)的发音字符串、用来设定发音的音的事件数据和用来控制上述应用软件14的事件数据等的用来进行语音合成的数据格式,为使由用户进行的手动输入变得容易而成为文本输入形式。

该脚本数据11中的数据格式的定义有语言依存性,可以由各种各样的语言所作的定义,但是在本实施例中,采取用日本语所作的定义。

用户短语合成辞典数据13和缺省合成辞典数据19通过用发音字符单元(例如,日本语的“あ”、“い”等)采样分析实际的人的声音,抽出8组共振

峰频率、共振峰强度 (formant level) 和音调作为参数, 并把该参数预先作成共振峰帧数据与发音字符单元对应起来, 相当于保存在发音字符单元的数据库。用户短语合成辞典数据 13 是构建在中间设备外的数据库, 用户可以对该数据库任意地登录共振峰帧数据, 因此, 可以经中间设备 API15 把用户短语合成辞典数据 13 的登录内容完全置换为缺省合成辞典数据 19 的保存内容。即, 可以把缺省合成辞典数据 19 的内容完全置换为用户短语合成辞典数据 13 的内容。另一方面, 缺省合成辞典数据 19 是构建在中间设备内的数据库。

作为用户短语合成辞典数据 13 和缺省合成辞典数据 19 最好分别具备男性音质用和女性音质用两类, 根据各帧的周期来改变语音再生装置 1 的语音输出, 但是被登录在用户短语合成辞典数据 13 和缺省合成辞典数据 19 内的共振峰帧数据的帧周期被设定为例如 20ms。

用户音色参数 12 和缺省音色参数 18 是控制语音再生装置 1 的语音输出中的音质的参数群, 即, 用户音色参数 12 和缺省音色参数 18 可以进行例如 8 组共振频率和共振峰强度的变更 (即, 来自登录在用户短语合成辞典数据 13 和缺省合成辞典数据 19 内的共振频率和共振峰强度的变化量的指定) 以及共振峰合成用的基本波形的指定, 这样, 就能够制作出各种各样的音色。

缺省音色参数 18 是作为缺省值预先设定在中间设备内的音色参数, 用户音色参数 12 是可以由用户任意作成的参数, 被设定保存在中间设备的外侧, 是经所述中间设备 API15 来扩展缺省音色参数 18 的内容的参数。

应用软件 14 是用来再生脚本数据 11 的软件。

中间设备 API15 构成由软件构成的应用软件 14 和由中间设备构成的变换器 16、驱动器 17、缺省音色参数 18 以及缺省合成辞典数据 19 间的接口。

变换器 16 解释脚本数据 11 并利用驱动器 17 最终变换为连结帧数据而形成的共振峰帧数据串。

驱动器 17 根据包含在脚本数据 11 内的发音字符串和缺省合成辞典数据 19 生成共振峰帧数据串, 并解释音色参数来加工该共振峰帧数据串。

音源 20 输出对应于变换器 16 的输出数据的合成发音信号, 把该合成发音信号输出到扬声器发出声音。

下面说明本实施例的语音再生装置 1 的技术特征。

用户音色参数 12 中包含有对任意发音单元分配用户短语合成辞典数据 13

存储的短语 ID 的参数。图 2 表示的是分配发音单元和短语 ID 的一例，这里，表示出音节 (mora) 与短语 ID 的分配关系。在日本语的情况下，所谓音节是“拍”的意思，例如相当于假名字符单元。

通过把短语 ID 分配给每个发音单元，用户音色参数 12 中所指定的发音单元不是参照缺省合成辞典数据 19 而是参照用户短语合成辞典数据 13 被规定下来。在用户音色参数 12 中，最好能从一个音色参数中指定任意的发音单元数。

上述用户音色参数 12 中的每个发音单元的短语 ID 分配是本实施例的一个示例，只要对应于发音单元，也可以采用其他方法。

然后来说明用户短语合成辞典数据 13 的细节，图 3 表示的是用户短语合成辞典数据 13 的内容例。用户短语合成辞典数据 13 存储由 8 组共振频率、共振峰强度和音调构成的共振峰帧数据，图 3 中的所谓“短语”是指例如日本语的“おはよう”等，具有一种意义或按音节统一的语句，该“短语”的定义规模无须特别规定，也可以是单词、音节和文章等任意规模的语句。

作为作成用户短语合成辞典数据 13 的工具，必须装载分析通常的声音文件（带有*.wav、*.aif 等扩展名的文件）来生成由 8 组共振频率、共振峰强度和音调构成的共振峰帧数据的分析工具。

脚本数据 11 中包含有指示音质变更的事件数据，但是可以用该事件数据来指定用户音色参数 12。

例如，作为脚本数据 11 的记述例，在使用日本语的平假名和字母数字字符的情况下，可以设定“TJK12 みなさん X10 あか”。该例中，“K”代表指定缺省音色参数 18 的事件数据，“X”代表指定用户音色参数 12 的事件数据。

“K12”是从多种缺省音色参数中指定某个特定缺省音色参数的代码，“X10”是从多种用户音色参数中指定图 2 所示的用户音色参数的代码。

在上述的例子中，被再生的合成语音为“みなさん こんにちは 鈴木です。”。这里，“みなさん”是参照缺省音色参数 18 和缺省合成辞典数据 19 再生的合成语音；“こんにちは”和“鈴木です”是参照用户音色参数 12 和用户短语合成辞典数据 13 再生的合成语音。即，语句“みなさん”是从缺省合成辞典数据 19 中读出并再生有关“み”、“な”、“さ”、“ん”4 个音素的共振峰帧数据的合成语音；语句“こんにちは”和“鈴木です”是从用户短语合成辞典数据 13 中读出并再生各自的短语单元的共振峰帧数据的合成语音。

在图2的示例中,虽然表示了“あ”、“い”、“か”等3个发音单元,但是只要是能够用文本标记的字符和符号,都不作特别规定。在上述的示例中,用接续“X10”的发音符号“あ”表示语句“こんにちは”来发音,而用发音符号“か”表示语句“鈴木です”来发音。因此,在上述发音例之后,在进行本来的“あ”的发音的情况下,只要插入使参照目标返回缺省合成辞典数据19的符号(例如,“X〇〇”、“〇”内插入规定的数字等)就可以。

下面参照图4来说明有关本实施例的语音再生装置1所用的音乐再生序列数据(SMAF: synthetic music mobile application format)的数据交换格式。图4表示SMAF文件格式,该SMAF是用音源表现音乐用的数据分配和用于相互利用的数据交换格式的一种,在便携式终端机(个人数字式助理(PDA),个人计算机,蜂窝电话等)中,是用来再生多媒体内容的数据格式的规格。

图4所示的数据交换格式的SMAF文件30把叫做块(chunk)的数据单元作为基本结构。块由固定长度(8字节)的首部和任意长度的主体部构成,首部分为4字节的块ID和4字节的块大小两部分。块ID被用作块的识别符,块大小表示主体部的长度。SMAF文件30其本身及其所包含的各种数据也构成为整个块结构。

如图4所示,SMAF文件30由内容信息块(contents info chunk)31、选择数据块(optional data chunk)32、乐谱音轨块(score track chunk)33以及HV块(HV chunk)36构成。

内容信息块31存储着有关SMAF文件30的各种管理信息,例如,存储内容的级别、种类、著作权信息、类型名、曲名、艺术家名、作词/作曲者名等信息。选择数据块32存储例如著作权信息、类型名、曲名、艺术家名、作词/作曲者名等信息。在SMAF文件30中,也不必设置操作数据块32。

乐谱音轨块33是存储发送到音源的乐曲的序列音轨的块,包含创建数据块(setup data chunk)34(选项)和序列数据块(sequence data chunk)35。

创建数据块34是存储音源的音色数据等的块,一并存储专用信息(exclusive message)的语句,作为专用信息例如有音色参数登录信息。

序列数据块35是存储实际演奏数据的块,把决定脚本数据11的再生定时的HV音符开(HV note-on,‘HV’表示人类的声音(human voice))与其他的序列事件混起来存储。这里,HV及其以外的乐曲事件由该HV的频道指定来区

别。

HV 块 36 包含 HV 创建数据块 (HV setup data chunk) 37 (选项)、HV 用户短语辞典块 (HV user phrase dictionary chunk) 38 (选项) 和 HV-S 块 39。

HV 创建数据块 37 存储 HV 用户音色参数和用来指定 HV 用的频道的信息, HV-S 块 39 存储 HV 脚本数据。

HV 用户短语辞典块 38 存储用户短语合成词典数据 13 的内容, 在存储于 HV 创建数据块 37 中的 HV 用户音色参数内必须要有用来分配图 2 所示的音节和短语 ID 的参数。

把图 4 所示的 SMAF 文件 30 适用于本实施例的音色参数中, 就能够与乐曲同步地再生出合成语音 (HV), 同时还能够再生用户短语合成词典数据 13 的内容。

然后参照图 5 说明用来作成图 1 所示的用户短语合成词典数据 13 和图 4 所示的 SMAF 文件 30 的工具即 HV 创作工具 (HV authoring tool)。图 5 是表示 HV 创作工具的功能和规格例的框图。

在作成 SMAF 文件 30 的情况下, HV 创作工具 42 读入预先由 MIDI (乐器数码接口; musical instrument digital interface) 序列发生器作成的 SMF 文件 (标准 MIDI 文件) 41 (包含决定 HV 的发音定时的音符开), 根据从 HV 脚本 UI (HV 脚本用户接口) 44 和 HV 语音编辑器 (HV voice editor) 45 得到的信息进行向 SMAF 文件 43 (相当于前述 SMAF 文件 30) 的变换处理。

HV 语音编辑器 45 是具有编辑包含在 HV 用户音色文件 48 内的 HV 用户音色参数 (相当于前述的用户音色参数 12) 的功能的编辑器。该 HV 语音编辑器 45 除编辑各种 HV 音色参数之外, 还能够对任意的音节分配用户短语。

HV 语音编辑器 45 的接口具有选择音节的菜单, 并具有对该音节分配任意声音文件 50 的功能。用 HV 语音编辑器 45 的接口被分配的声音文件 50 由波形分析器 46 进行分析, 由此来生成 8 组共振频率、共振峰强度和音调构成的共振峰帧数据。该共振峰帧数据可以作为个别文件 (即, HV 用户音色文件 48、HV 用户合成辞典文件 49) 输出输入。

HV 脚本 UI44 可以直接编辑 HV 脚本数据, 该 HV 脚本数据也可以作为个别文件 (即, HV 脚本文件 47) 输入输出。另外, 涉及本实施例的 HV 创作工具

40, 也可以仅由上述的 HV 创作工具 42、HV 脚本 UI44、HV 语音编辑器 45 和波形分析器 46 构成。

下面参照图 6 说明把本实施例的语音再生装置 1 适用于便携式通信终端的示例, 图 6 是表示具备语音再生装置 1 的便携式通信终端 60 的构成的框图。

便携式通信终端 60 例如是相当于便携式电话机的设备, 设置有 CPU61、ROM62、RAM63、显示部 64、振动器 65、输入部 66、通信部 67、天线 68、语音处理部 69、音源 70、扬声器 71 以及总线 72。CPU61 进行便携式通信终端 60 总体的控制, ROM62 存储各种通信控制程序和用来再生乐曲的程序等控制程序, 同时存储各种常数数据等。

RAM63 被用作工作区, 同时存储乐曲文件和各种应用程序。显示部 64 例如由液晶显示装置 (LCD) 构成, 振动器 65 在便携式电话机有来话呼叫时振动。输入部 66 由多个键等操作件构成, 这些操作件根据用户的操作指示用户音色参数、用户短语合成辞典数据和 HV 脚本数据的登录处理。通信部 67 由调制解调器等构成, 连接在天线上。

语音处理部 69 连接在送话器和受话扬声器 (例如, 扩音器和耳机; e. g., microphone and earphone) 上, 具有为了进行通话而对语音信号进行编码和译码的功能。音源 70 根据存储在 RAM63 等的乐曲文件进行乐曲的再生, 同时再生语音信号并输出到扬声器 71。总线 72 是用来进行由 CPU61、ROM62、RAM63、显示部 64、振动器 65、输入部 66、通信部 67、语音处理部 69 和音源 70 构成的便携式电话机的各构成要素之间的数据传送的传送路径。

通信部 67 可以从规定的内容服务器 (Contents Server) 等下载 HV 脚本文件或图 4 所示的 SMAF 文件 30, 并存储在 RAM63 内。在 ROM62 中存储有图 1 所示的语音再生装置 1 的应用程序 14 和中间设备的程序。CPU61 读出并启动该应用程序 14 和中间设备的程序。CPU61 解释存储在 RAM63 内的 HV 脚本数据, 并生成共振峰帧数据, 把该共振峰帧数据送到音源 70。

然后说明本实施例的语音再生装置 1 的动作。首先, 说明用户短语合成辞典 13 的作成方法。图 7 是表示用户短语合成辞典 13 的作成方法的流程图。

首先, 在步骤 S1, 用图 5 所示的 HV 创作工具 42 选择参照用户短语合成辞典 13 的 HV 音色, 启动 HV 语音编辑器 45。然后用 HV 语音编辑器 45 选择使用的音节, 并装上声音文件。这样, 在步骤 S2, HV 语音编辑器 45 生成并输出

用户短语辞典数据(相当于 HV 用户合成辞典文件 49)。

然后,用 HV 语音编辑器 45 编辑 HV 音色参数,接着,在步骤 S3, HV 语音编辑器 45 生成并输出用户音色参数(相当于用户音色文件 48)。

然后,用 HV 脚本 UI44 在 HV 脚本数据中记述指定相应的 HV 音色的音质变更事件,由此来记述想要再生的音节。接着,在步骤 S4, HV 脚本 UI44 生成并输出 HV 数据(相当于 HV 脚本文件 47)。

然后,参照图 8 说明语音再生装置 1 中的用户短语合成辞典数据 13 的再生动作。图 8 是表示语音再生装置 1 中的用户短语合成辞典数据 13 的再生动作的流程图。

首先,在步骤 S11,把用户音色参数 12 和用户短语合成辞典数据 13 登录到语音再生装置 1 的中间设备中。然后,把脚本数据 11 登录在语音再生装置 1 的中间设备中,在步骤 S12 开始 HV 脚本数据的再生。

再生时,在步骤 S13,监视指定用户音色参数 12 的音质变更事件(X 事件)是否包含在脚本数据 11 中。

在步骤 S13,如果发现了音质变更事件,从用户音色参数 12 中探寻给音节分配的短语 ID,并从用户短语合成辞典数据 13 中读出对应于该短语 ID 的数据,然后,在步骤 S14,把 HV 驱动器管理的缺省合成辞典数据 19 内的相应的音节的辞典数据置换为用户短语合成辞典数据 13。步骤 S14 的替换处理也可以在再生 HV 脚本数据之前进行。

步骤 S14 结束之后,或者在步骤 S13 未发现音质变更事件的情况下,流程进到步骤 S15,变换器 16 解释脚本数据 11(当进行了步骤 S14 的处理时,该步骤 S14 的交换处理后的脚本数据)的音节,最后用 HV 驱动器变换为共振峰帧串数据。

在步骤 S16,通过音源 20 再生由步骤 S15 变换得到的数据。

此后,进到步骤 S17,判定脚本数据 11 的再生是否已经结束,在未结束的情况下,返回到步骤 S13,另一方面,如果结束了,就结束图 8 所示的用户短语合成辞典数据 13 的再生处理。

然后,参照图 9 说明图 4 所示的 SMAF 文件 30 的作成方法。图 9 是表示 SMAF 文件 30 的作成方法的流程图。

首先,按照图 7 所示的步骤,作用户短语合成辞典数据 13、用户音色

参数 12 和脚本数据 11 (参照步骤 S21)。

然后, 在步骤 S22, 作成包含控制乐曲数据和 HV 脚本数据的发音的事件的 SMF 文件 41。

接下来, 将 SMF 文件 41 输入到图 5 所示的 HV 创作工具 42, 用该 HV 创作工具 42 把 SMF 文件 41 变换为 SMAF 文件 43 (相当于前述的 SMAF 文件 30) (参照步骤 S23)。

然后, 把在前述步骤 S21 作成的用户音色参数 12 输入到图 4 所示的 SMAF 文件 30 的 HV 块 36 内的 HV 创建数据块 37, 另外, 将在步骤 S21 作成的用户短语合成词典数据 13 输入到 SMAF 文件 30 的 HV 块 36 内的 HV 用户短语辞典块 38, 就这样生成并输出输入 SMAF 文件 30 (参照步骤 S24)。

接下来, 参照图 10 说明 SMAF 文件 30 的再生处理, 图 10 是 SMAF 文件 30 的再生处理流程图。

首先, 在步骤 S31, 把 SMAF 文件 30 登录到图 1 所示的语音再生装置 1 的中间设备内。这里, 通常语音再生装置 1 把 SMAF 文件 30 内的乐曲数据部分登录在中间设备的乐曲再生部中, 进行再生准备。

在步骤 S32, 语音再生装置 1 判定 HV 块 36 是否包含在 SMAF 文件 30 内。

在步骤 S32 的判定结果为“是”的情况下, 流程进入到步骤 S33, 语音再生装置 1 解释 HV 块 36 的内容。

在步骤 S34, 语音再生装置 1 进行用户音色参数的登录、用户短语辞典数据的登录和 HV 脚本数据的登录。

在步骤 S32 的判定结果为“否”的情况下, 或在步骤 S34 中的登录处理结束了的情况下, 进到步骤 S35, 语音再生装置 1 进行该乐曲再生部内的块的解释。

然后, 语音再生装置 1 对应于“开始”信号, 开始解释序列数据块 35 内的序列数据 (即, 实际演奏数据), 由此来进行乐曲再生 (参照步骤 S36)。

在上述的乐曲再生中, 语音再生装置 1 按顺序解释包含在序列数据内的事件, 在该过程中, 判定各事件是否相当于 HV 音符开 (参照步骤 S37)。

在步骤 S37 的判定结果为“是”的情况下, 流程进到步骤 S38, 语音再生装置 1 开始再生由 HV 音符开指定的 HV 块的 HV 脚本数据。

步骤 S38 结束之后, 语音再生装置 1 进行图 8 所示的用户短语合成辞典数

据的再生处理,即,在步骤 S38 中的 HV 脚本数据的再生中,语音再生装置 1 判定是否存在指定用户音色参数 12 的音质变更事件(X 事件)(参照步骤 S39)。

存在上述音质变更事件的情况下,即,步骤 S39 的判定结果为“是”的情况下,流程进到步骤 S40,从用户音色参数 12 中探寻给音节分配的短语 ID,并从用户短语合成辞典数据 13 读出对应于短语 ID 的数据,把 HV 驱动器管理的缺省合成辞典数据 19 内相应的音节的辞典数据置换为用户短语合成辞典数据。该步骤 S40 的替换处理也可以在再生 HV 脚本数据之前进行。

步骤 S40 结束之后,或者在步骤 S39 未发现音质变更事件的情况下,流程进到步骤 S41,变换器 16 解释脚本数据 11 的音节,用 HV 驱动器最终变换为共振峰帧串数据。

然后流程进到步骤 S42,在音源 20 的 HV 再生部把在步骤 S41 中被变换的数据再生出来。

此后,流程进到步骤 S43,语音再生装置 1 判定乐曲的再生是否已经结束。乐曲再生结束了的情况下,结束 SMAF 文件 30 的再生处理,另一方面,在乐曲再生未结束的情况下,流程返回到步骤 S37。

在步骤 S37,所述序列数据中的事件不是 HV 音符开的情况下,语音再生装置 1 就把该事件认作乐曲数据的一部分,并变换为音源再生事件数据(参照步骤 S44)。

然后流程进到步骤 S45,语音再生装置 1 把在步骤 S44 被变换的数据在音源 20 的乐曲再生部再生出来。

如上所述,本实施例采用了使用 FM 音源的共振峰合成进行的语音再生方式,有如下 3 个优点。

(1) 能够分配用户喜好的短语,即,不依存于固定辞典,就能够以更近似于喜好的音色的音色进行语音再生。

(2) 因为把缺省合成辞典数据 19 的一部分置换为用户短语合成辞典数据 13,所以能够避免在语音再生装置 1 中过大地增加数据容量。由于能够把缺省合成辞典数据 19 的一部分置换为任意的短语,所以能够实现按短语单元的发音,并能够消除现有的按发音单元的合成语音中产生的各发音单元之间的连接点中引起的听觉上的不和谐的感觉。

(3) 由于能够在 HV 脚本数据中指定任意的短语,所以可以并用音节单元

的语音合成和短语单元的语音发音。

另外,按照本实施例,与再生预先对短语采样构成的波形数据的方法相比,能够实现按照共振峰强度的音色变化。虽然本实施例中的数据大小和品质都依存于帧率,但是与现有的由采样波形数据进行的方法相比,能够用远少的数据容量实现高品质的语音再生。因此,可以容易地把例如本实施例的语音再生装置1组装在便携式电话机等便携式通信终端中,因此,能够以高品质的语音再生电子邮件等的內容。

图11是表示本发明的第二实施例的语音·乐曲再生装置的构成的框图。这里,HV-脚本(即HV脚本数据)是相当于定义用来再生语音的格式的文件,是定义用来进行由包含韵律符号(即,指定音调等发音形态的符号)的发音字符串、发音的音的设定和对再生应用等的信息构成的语音合成的数据的文件,为了使由用户的作成变得更容易而用文本输入来作成。

HV-脚本由文本编辑器等应用软件来读入,只要能由文本进行编辑的文件形式来记述就行,作为一例,可列举出用文本编辑器作成的文本文件。在HV-脚本中有语言依存性,可以用各种语言定义,在本实施例中,HV-脚本用日语来定义。

符号101表示HV-脚本播放器(HV-Script player),用来控制HV-脚本的再生或停止等。这里,HV-脚本被登录在HV-脚本播放器101中并接受到其再生指示的情况下,HV-脚本播放器101开始解释该HV-脚本。然后,根据HV-脚本中记述的事件的种类对HV驱动器102、波形再生播放器104和短语再生播放器107中的某一个进行根据该事件的处理。

HV驱动器102从未图示的ROM(智读存储器;read-only memory)中读出并参照合成辞典数据,人的语音具有依存于人体的构造(例如声带或口腔等的形状)的规定的共振峰(即,固有频谱),合成辞典数据把有关语音的共振峰的参数与发音文字对应起来进行保存。合成辞典数据相当于把按发音文字单元(例如,日语的“あ”、“い”等音素单元)对实际的声音进行采样及分析的结果得到的参数作为共振峰帧数据预先按发音文字单元存储起来的数据库。

例如,前述的CSM(Composite Sinusoidal Modeling:复合正弦波模型)语音合成方式的情况下,合成辞典数据把8组共振频率、共振峰强度和音调等作为参数保存起来。这样的语音合成方式与采样语音作成的波形数据的再生方

式相比,具有数据量非常小的优点。而且,在合成辞典数据中,还可以保存控制被再生的语音的音质的参数(例如,用来进行8组共振频率和共振峰强度的变更指定的参数等)。

HV驱动器102解释包含HV-脚本中的韵律符号的发音字符串等,并用合成辞典数据变换为共振峰帧串之后,输出到HV音源103。HV音源103根据从HV驱动器102输出的共振峰帧串生成发音信号并将其输出到加法器110。

波形再生播放器104进行预先采样语音或乐曲以及模拟音等的波形数据的再生或停止等。符号105代表波形数据用RAM(波形数据随机存取存储器; waveform data random-access memory),预先存储有缺省波形数据。用户可以经由登录API(登录应用程序接口; registration application program interface)113把用户数据用RAM112中的用户波形数据存储在波形数据用RAM105内。如果波形再生播放器104接受来自HV-脚本播放器101的再生指示,则从波形数据用RAM105读出波形数据并输出到波形再生器106。波形再生器106根据从波形再生播放器104输出的波形数据生成发音信号并输出到加法器110。被采样的波形数据不限于PCM(脉冲编码调制; pulse-code modulation)方式,例如也可以采取MP3(移动图象专家组层面3; moving picture experts group layer 3)方式的语音压缩格式。

短语再生播放器107进行乐曲短句数据(或乐曲数据)的再生或停止等。乐曲短句数据是SMF格式形式,由代表发音的声音的音调和音量等的音符信息和表示发音的声音的发音时间的时间信息构成。符号108代表乐曲短句数据用RAM,预先存储有缺省乐曲短句数据。用户可以经由登录API把用户数据用RAM112中的用户乐曲短句数据存储在乐曲短句数据用RAM108内。

短语再生播放器107一旦接受来自HV-脚本播放器101的再生指示,就从乐曲短句数据用RAM108读出乐曲短句数据,并进行乐曲短句数据中的音符信息的时间管理,根据乐曲短句数据中记述的时间信息把音符信息输出到短语音源109。短语音源109根据短语再生播放器107输出的音符信息生成乐音信号,并输出到加法器110。作为短语音源109可以采用FM方式或PCM方式等,但是如果具有乐曲短句数据的再生功能,就不必限定该音源的方式。

加法器110把从HV音源103输出的发音信号、从波形再生器106输出的语音信号和从短语音源109输出的乐音信号合成起来,并把该合成信号输出到

扬声器 111。扬声器 111 根据加法器 110 的合成信号发出语音和/或乐音。

也可以在 HV 驱动器 102、波形再生播放器 104 和短语再生播放器 107 中同时进行处理，由此使分别基于发音信号、语音信号和乐音信号（另外，也可以把语音信号和乐音信号统称为“声音信号”）的语音和乐曲同时发出声音来。或者，也可以用 HV-脚本播放器 101 管理 HV 驱动器 102、波形再生播放器 104 和短语再生播放器 107 的处理定时，同时再生基于各自的处理的语音和乐曲。在本实施例中，禁止由 HV 驱动器 102、波形再生播放器 104 和短语再生播放器 107 的同时处理。另外，在图 11 中，为方便说明，分别设置各自的 RAM 作为波形数据用 RAM105、乐曲短句数据用 RAM108 和用户数据用 RAM112，但是也可以把这些功能分配到单一的 RAM 内的不同的存储区域。

图 12 示出用来再生 HV-脚本记述的波形数据或乐曲短句数据（以下统称“声音数据”）的事件的定义例。事件的开头文字“D”的含义是缺省定义，“O”的含义是用户定义。作为各事件的类别，分配为波形或短语。把波形数据用 RAM105 预先存储的缺省波形数据或乐曲短句数据用 RAM108 预先存储的缺省乐曲短句数据分配到缺省定义（D0～D63）中，缺省定义中可以分配 64 个缺省波形数据和缺省乐曲短句数据。把用户任意作成的采样波形数据或乐曲短句数据分配到用户定义（O0～O63）中，用户定义中可以分配 64 个采样波形数据或乐曲短句数据。

图 12 所示的类别是波形数据的事件和表示与该事件代表的波形数据的关系的数据被预先存储在波形数据用 RAM105 内。另外，类别是短语的事件和表示与该事件代表的乐曲短句数据的关系的数据被预先存储在乐曲短句数据用 RAM108 内。在用户进行过用户数据用 RAM112 中的波形数据或乐曲短句数据的登录的情况下，把这些数据更新。

作为 HV-脚本例如记述为“TJK12 みなさん O0 です。D20”，开头记述的“TJK12”内的“T”是表示 HV-脚本的开始符号，“J”指定国家文字代码，这里表示 HV-脚本用日语来记述。“K12”是设定音质的符号，表示指定第 12 种音质。另外，“みなさん”和“です”用 HV 驱动器 102 来解释，从扬声器 111 发出“みなさん”和“です”那样的日语的语音。在“みなさん”和“です”那样的发音字符串中包含有语调（或强弱）等表示发音状态的韵律符号的情况下，发出加了语调（或加了强弱的）的语音。

在用户事件“○0”中，例如登录有对发出“铃木”声音的语音进行采样的波形数据。该用户事件“○0”用波形再生播放器 104 来解释，由此，从扬声器 111 发出“铃木”的语音。另外，在用户事件“○20”中登录有例如欢快的短的乐曲短句数据。该用户事件“○20”用短语再生播放器 107 解释，由此，从扬声器 111 发出欢快的乐曲音。这时，再生语音成为“みなさん鈴木です”（乐曲短句再生时），只有“铃木”部分再生波形数据。由波形数据的再生产生的语音的发音与由“みなさん”或“です”那样的发音单元的语音合成再生的音势相比，发音单元的连接点的再生就更加自然。另外，把叫做“铃木”的语句的发音作为特色的波形的再生就能够让用户有效地听到再生的语音。如上所述，由 HV-脚本记述指定波形数据或乐曲短句数据的再生的事件，就能够任意指定波形数据或乐曲短句数据的再生定时。有关 HV-脚本的记述的设定是所谓设计事项，不限于上述的方法。

然后，用图 13 的流程图来说明涉及本实施例的语音・乐曲再生装置的动作。首先，用户用文本编辑器作成 HV-脚本，并登录在所述 HV-脚本播放器 101 内（参照步骤 S101）。这时，如果存在由用户定义产生的波形数据或乐曲短句数据，登录 API113 就从用户数据用 RAM112 读入波形数据或乐曲短句数据。登录 API113 把波形数据存储在波形数据用 RAM105 内，而把乐曲短句数据存储在乐曲短句数据用 RAM108 内。

一旦用户作出开始指示（步骤 S103），HV-脚本播放器 101 就开始解释 HV-脚本（参照步骤 S102）。HV-脚本播放器 101 判定 HV-脚本中是否包含从“D”或“○”开始的事件（步骤 S104）；在有从“D”或“○”开始的事件的输入时，判定其类别是否是波形数据（步骤 S105）。如果该事件的类别是波形数据，则 HV-脚本播放器 101 对波形再生播放器 104 指示该处理，波形再生播放器 104 从波形数据用 RAM105 读出接续“D”或“○”的号码的波形数据，并输出到波形再生器 106（步骤 S106）。波形再生器 106 根据该波形数据生成语音信号，经加法器 110 输出到扬声器 111（步骤 S107）。就这样，扬声器 111 发出相应的语音。

另外，在步骤 S105 中事件类别不是波形数据的情况下，流程进到步骤 S108，HV-脚本播放器 101 判定事件的类别是否是乐曲短句数据。在事件的类别是乐曲短句数据的情况下，HV-脚本播放器 101 对短语再生播放器 107 指示

该处理。短语再生播放器 107 从乐曲短句数据用 RAM108 读出接续“D”或“O”的号码的波形数据,根据该乐曲短句数据中的时间信息把乐曲短句数据中的音符信息输出到短语音源 109 (参照步骤 S109)。短语音源 109 根据该音符信息生成乐音信号,经加法器 110 输出到扬声器 111 (步骤 S110)。就这样,扬声器 111 发出乐曲音。另外,在步骤 S108,事件类别被判定为不是乐曲短句数据的情况下,本实施例的语音・乐曲再生装置认为是不能处理的事件,流程进到步骤 S113。

在步骤 S104, HV-脚本中未记述从“D”开始或从“O”开始的事件的情况下, HV-脚本播放器 101 对 HV 驱动器 102 指示其处理。HV 驱动器 102 用合成辞典数据把字符串变换为共振峰帧串,输出到 HV 音源 103 (参照步骤 S111)。HV 音源 103 根据该共振峰帧串生成发音信号,并经加法器 110 输出到扬声器 111 (参照步骤 S112)。就这样,扬声器 111 发出相应的语音。

每当有关事件的处理结束时, HV-脚本播放器 101 判定到 HV-脚本的最后记述为止是否结束了解释 (步骤 S113)。在还剩余有应解释的记述时,返回到步骤 S104, 另一方面,在 HV-脚本的全部记述都解释完了的情况下,结束图 13 所示的语音・乐曲再生处理。

作为本实施例的 HV-脚本的记述例所示出的“TJK12 みなさん〇〇です。D20”的情况下,用事件“〇〇”定义的波形数据的发音结束之后,必须发出下一个语句“です”的语音。例如,在 HV-脚本播放器 101 进行波形数据 (或乐曲短句数据) 的事件的解释的情况下,暂时缓期其下一个事件的再生,在结束由波形再生播放器 104 (或短语再生播放器 107) 进行的发音时,从该波形再生播放器 104 对 HV-脚本播放器 101 输出表示发音结束的信号。

另外,在允许由 HV 驱动器 102、波形再生播放器 104 和短语再生播放器 107 同时进行再生处理的情况下,也可以用 HV-脚本的记述来控制它们的再生处理。例如,在 HV-脚本记述着“TJK12 みなさん〇〇 3 です。D20”的情况下,用接续“〇〇”的“ ” (空格) 和“3”表示设置规定的无音期间的事件,使得在发出“〇〇”表示的语句“铃木”的语音间由 HV 驱动器 102 再生的语音为无声的方式控制。在 HV-脚本记述有“TJK12 こんにちは。D20 みなさん〇〇 3 です。”的情况下,由“D20”指定的乐曲和叫做“みなさん鈴木です”的语音同时发音。

图 14 是具备本实施例的语音·乐曲再生装置的便携式电话机的构成框图。这里，标号 141 代表控制便携式电话机的各部的 CPU。标号 142 代表发送接收数据用的天线。标号 143 代表通信部，把发送用数据调制后输出到天线 142，同时对由天线接收到的接收用数据进行解调。标号 144 代表语音处理部，在便携式电话机进行通话时把从通信部 143 输出的通话对方的语音数据变换为语音信号输出到近耳扬声器（或耳机，未图示），同时把从送话器（未图示）输入的语音信号变换为语音数据输出到通信部 143。

标号 145 代表音源，具有与图 11 所示的 HV 音源 103、波形再生器 106 和短语音源 109 同样的功能。标号 146 代表扬声器，发出所希望的语音或乐音。标号 147 代表由用户操作的操作部。标号 148 代表存储由 HV-脚本或用户定义的波形数据和乐曲短句数据等的 RAM。标号 149 代表存储 CPU141 执行的程序、以及合成辞典数据、缺省波形数据、缺省乐曲短句数据等的 ROM。标号 150 代表显示部，把用户进行的操作结果或便携式电话机的状态等显示在画面上。标号 151 代表振动器，在便携式电话机有来话呼叫时接受来自 CPU141 的指示，产生振动。上述的各功能框经总线 B 相互连接起来。

该便携式电话机具有从语音生成波形数据的功能，把从送话器输入的语音送到语音处理部 144，并变换为波形数据，把该波形数据存储于 RAM148 内。在通过通信部 143 从 WEB 服务器下载乐曲短句数据时，就把该乐曲短句数据存储于 RAM148 内。

CPU141 按照存储于 ROM149 内的程序，进行与图 11 所示的 HV-脚本播放器 101、HV 驱动器 102、波形再生播放器 104 和短语再生播放器 107 同样的处理。另外，CPU141 还对从 RAM148 读出的记述在 HV-脚本内的事件进行解释。在事件表示由语音合成进行发音的情况下，CPU141 从 ROM149 读出并参照合成辞典数据，把记述在 HV-脚本内的字符串变换为共振峰帧串并将其输出到音源 145。

在事件表示波形数据的再生的情况下，CPU141 从 RAM148 或 ROM149 中读出 HV-脚本中的接续“D”或“O”的号码的波形数据并输出到音源 145。在事件表示乐曲短句的再生的情况下，CPU141 从 RAM148 或 ROM149 中读出 HV-脚本中的接续“D”或“O”的号码的乐曲短句数据，并根据该乐曲短句数据中的时间信息把乐曲短句数据中的音符信息输出到音源 145。

音源 145 根据从 CPU141 输出的共振峰帧串生成合成发音信号并输出到扬

声器 146。并根据从 CPU141 输出的波形数据生成语音信号并输出到扬声器 146。进一步, 还根据从 CPU141 输出的乐曲短句数据生成乐音信号并输出到扬声器 146。扬声器 146 根据合成发音信号、语音信号或乐音信号适宜地发出语音或乐音。

一旦用户操作操作部 147 来启动对应于文本编辑的软件, 则用户就能认识显示在显示部 150 画面上的内容的同时作成 HV-脚本, 能够把这样作成的 HV-脚本存储在 RAM148 中。

另外, 可以把由用户作成的 HV-脚本应用于来电呼叫铃声。这时, 在便携式电话机有来电呼叫时使用 HV-脚本的事宜作为设定信息, 预先存储在 RAM148 内。即, 通信部 143 经天线 142 接收从其他便携式电话机发送的呼叫信息时, 通信部 143 对 CPU141 通知有来电。接受了来电通知的 CPU141 从 RAM148 读出设定信息, 再从 RAM148 读出该设定信息表示的 HV-脚本, 并开始其解释。以后的处理如前所述。即, 扬声器 146 按照 HV-脚本内记述的事件的类别发出语音或乐音。

另外, 用户还可以把 HV-脚本添加在电子邮件内, 发送到其他终端。另外, CPU141 也可以按照 HV-脚本形式解释电子邮件的文本本身, 接受用户的指示后, 根据电子邮件中的记述向语音处理部 144 输出该 HV-脚本的再生指示。HV-脚本播放器 101、HV 驱动器 102、波形再生播放器 104 和短语再生播放器 107 的全部功能不必由 CPU141 来负担。例如也可以由音源 145 来负担上述功能的任意项功能。

本实施例的适用对象并不局限于便携式电话机 (cellular phone), 例如适用于 PHS (personal handphone system: (个人手持电话系统) 日本国注册商标) 或 PDA (个人数字助理) 等便携式终端, 进行前述的语音和乐曲再生。

另外, 作为本实施例的灵活应用例, 在便携式电话机等便携式移动终端中可以输入由用户作成的 HV-脚本, 由此, 一般用户不仅可以容易地作成语音合成用的文字, 而且能够容易地作成用来再生定型的采样波形数据或乐曲短句数据的 HV-脚本。另外, 在发送和接收用的便携式终端中具备本实施例的语音・乐曲再生装置的情况下, 用户能够操作便携式移动终端, 把 HV-脚本添加在电子邮件内进行发送接收。这样, 用接收方的便携式移动终端接收到的电子邮件不仅可以适宜地再生语音合成用的文字, 而且可以适宜地再生定型的采样数据或

乐曲短句数据。另外，还可以将使用 HV-脚本的语音和乐曲的再生作为来电呼叫铃声利用。

然后，参照图 15 和图 16 说明本发明的第三实施例的语音・乐曲再生装置。第三实施例是将前述的第一实施例与第二实施例组合起来构成的，中间设备实施 HV 再生、波形再生和乐曲短句再生，音源根据这 3 种数据产生合成发音信号，把来自 3 个系统的信号合成起来输出到扬声器。

这里，图 15 是以图 1 的结构为基本，将图 11 的部分结构组合起来的构成，标号 211~219 对应于图 1 的标号 11~19，标号 303~313 对应于图 11 的标号 103~113。即，图 11 的用户数据用 RAM 通过用户数据 API 连接在图 1 的中间设备上，在中间设备内，追加了图 11 的波形再生播放器和短语再生播放器，该波形再生播放器和短语再生播放器分别与波形数据用 RAM 和乐曲短句数据用 RAM 连接。另外，应用软件兼有作为图 11 的 HV 脚本播放器的功能，HV 脚本通过中间设备 API 根据记述在 HV-脚本的事件的种类向 HV 变换、波形再生播放器、乐曲短句播放器的任一个指示处理。另外，音源兼具有图 11 的 HV 音源、波形发生器、以及乐曲短句音源的三种功能，他们的各输出信号通过加法器被合成并在扬声器发音。另外，图 15 所示的各构成要素的动作与图 1 以及图 11 所示的对应构成要素的动作相同，故省略其详细说明。

图 16 是表示图 15 所示的语音・乐器再生装置的动作的流程图。其以图 13 所示的流程图为基本，追加了图 8 所示的流程图的一部分，标号 S211~S216 对应于图 8 的标号 S11~S16，标号 S304~S310，S312，S313 对应于图 13 的标号 S104~S110，S112，S113。即，图 13 的步骤 S104 中的判断结果为“否”的情况下，执行与图 8 的步骤 S13、S14、S15 相同的处理，接着在步骤 S312 中执行 HV 音源再生处理。就这样，通过输入一个 HV-脚本可以进行根据 HV 音源的语音的再生、根据波形再生播放器的波形数据的再生、以及可以进行基于根据乐曲短句再生播放器的音符信息的乐曲短句的再生。另外，图 16 所示的各步骤的处理与图 8 和图 13 相同，所以省略起详细说明。

最后，说明前述的实施例中所使用的韵律符号。例如，记述在 HV-脚本内的“は[^]3じま\$り[^]ま\$5し>10た。”就是在“はじまりました”（即发音的字符串）内附加规定的音调（イントネーション）而进行语音合成起来的，这里，“[^]”、“\$”、“>”等相当于韵律符号。对该韵律符号后面的文字（在韵律符

号后紧接数值的情况下，接续在该数值后面的文字）附加规定的昂扬（语调）。

具体地说，“^”表示发音中音调提高，“\$”表示发音中音调降低，“>”表示发音中音量下降，根据这些符号进行语音合成。在韵律符号后紧接数值的情况下，该数值用来指定附加的语调的变化量。例如，语句“は^ 3 じま”的情况下，“は”按标准音调和音量发音，在发音中把“じ”提高音调“3”的量，后面的“ま”按被提高的音调发音。

这样，在包含在发音的语言内的文字中附加规定的语调（或音调）的情况下，在该文字跟前记述上述的韵律符号（还有表示音调变化量的数值）。上述的韵律符号用来控制发音中的音调或音量，但是并不限于此，例如也可以使用控制音质或速度的符号。把这样的符号附加在 HV-脚本中就能够很合适地体现出音调等发音状态。

另外，本发明并不限于上述的实施例，发明范围内的变更均被包含在本发明之中。

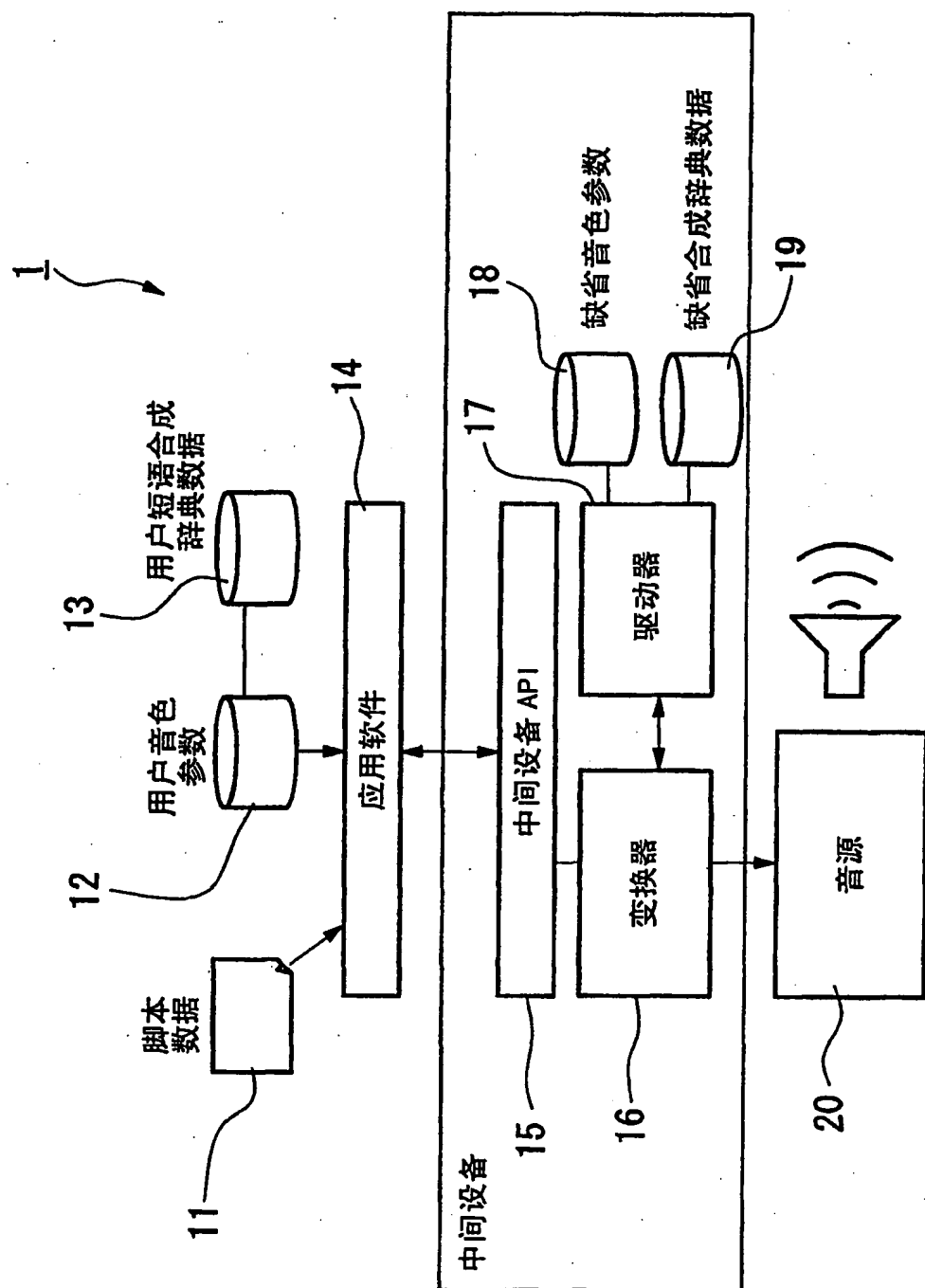


图 1

音节与短语 ID 的分配

发音单元	短语 ID
“あ”	1
“い”	2
“か”	10

图 2

短语 ID	数据
1	对短语“こんにちは”的 8 组共振频率、共振峰强度和音调
2	对短句“い”的 8 组共振频率、共振峰强度和音调
10	对短语“すずきです”的 8 组共振频率、共振峰强度和音调

图 3

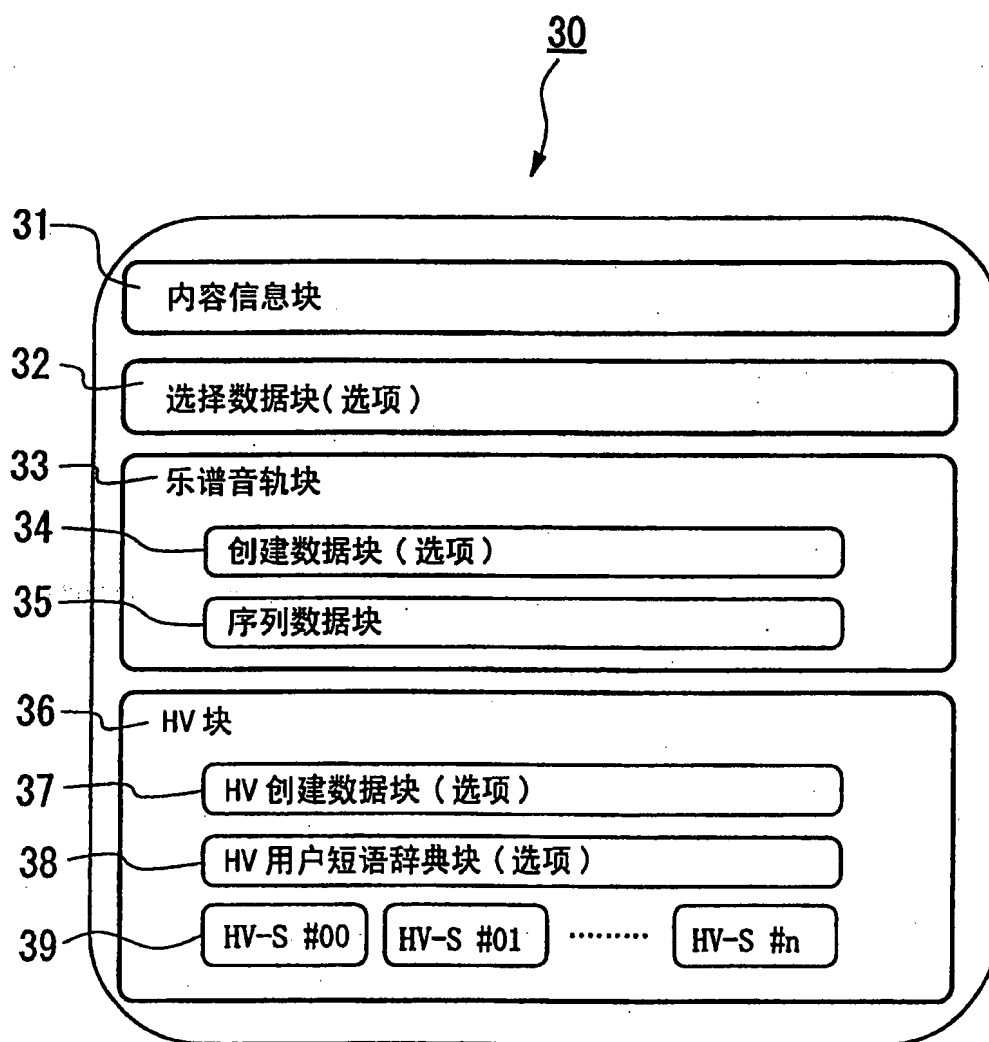


图 4

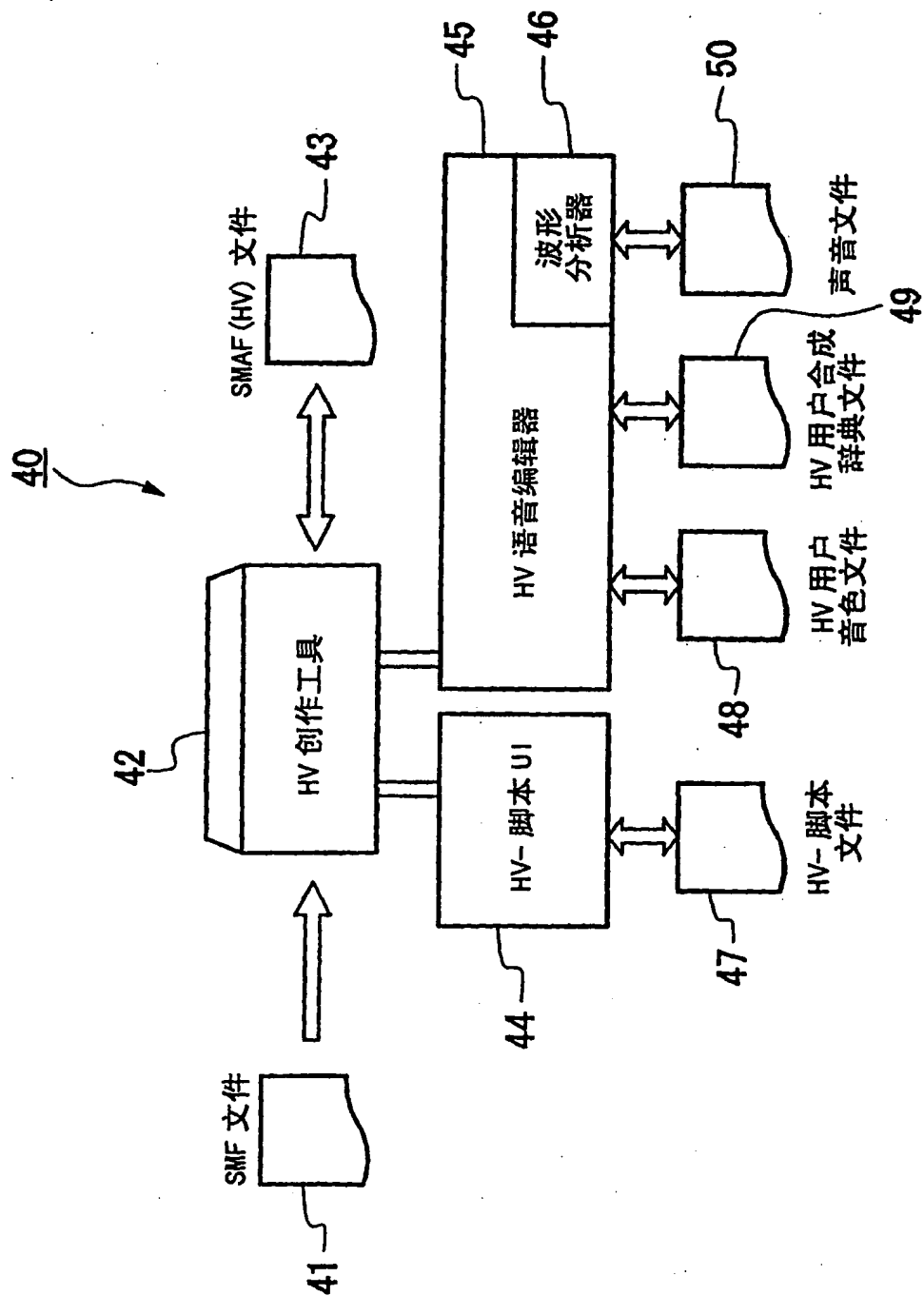


图 5

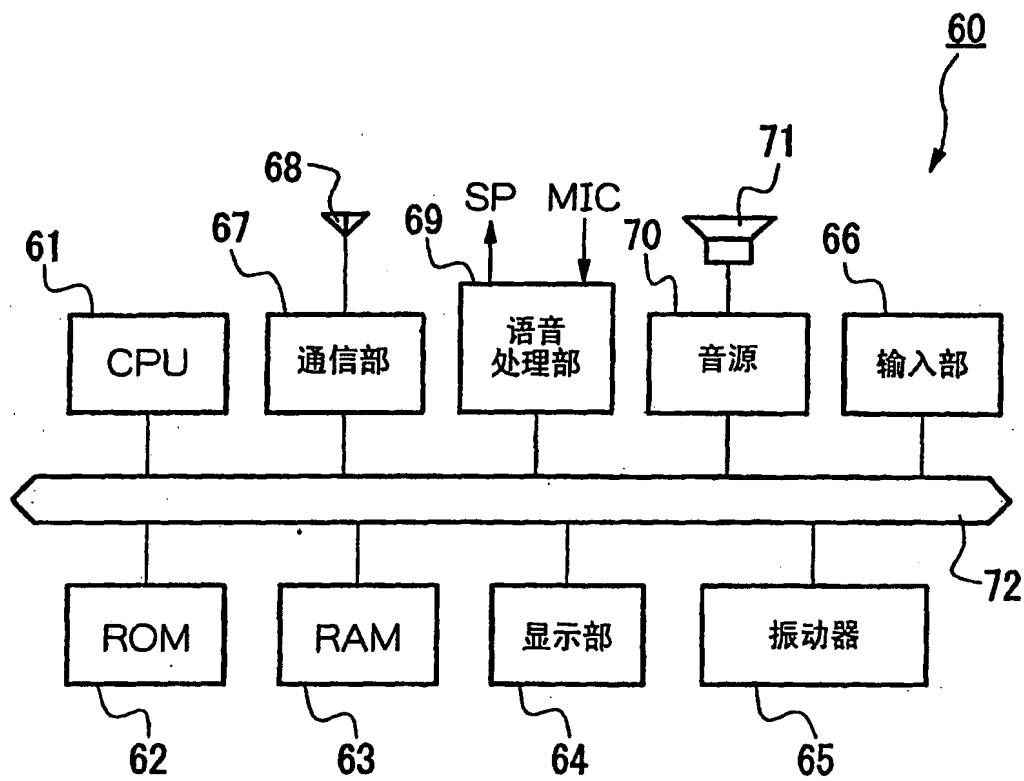


图 6

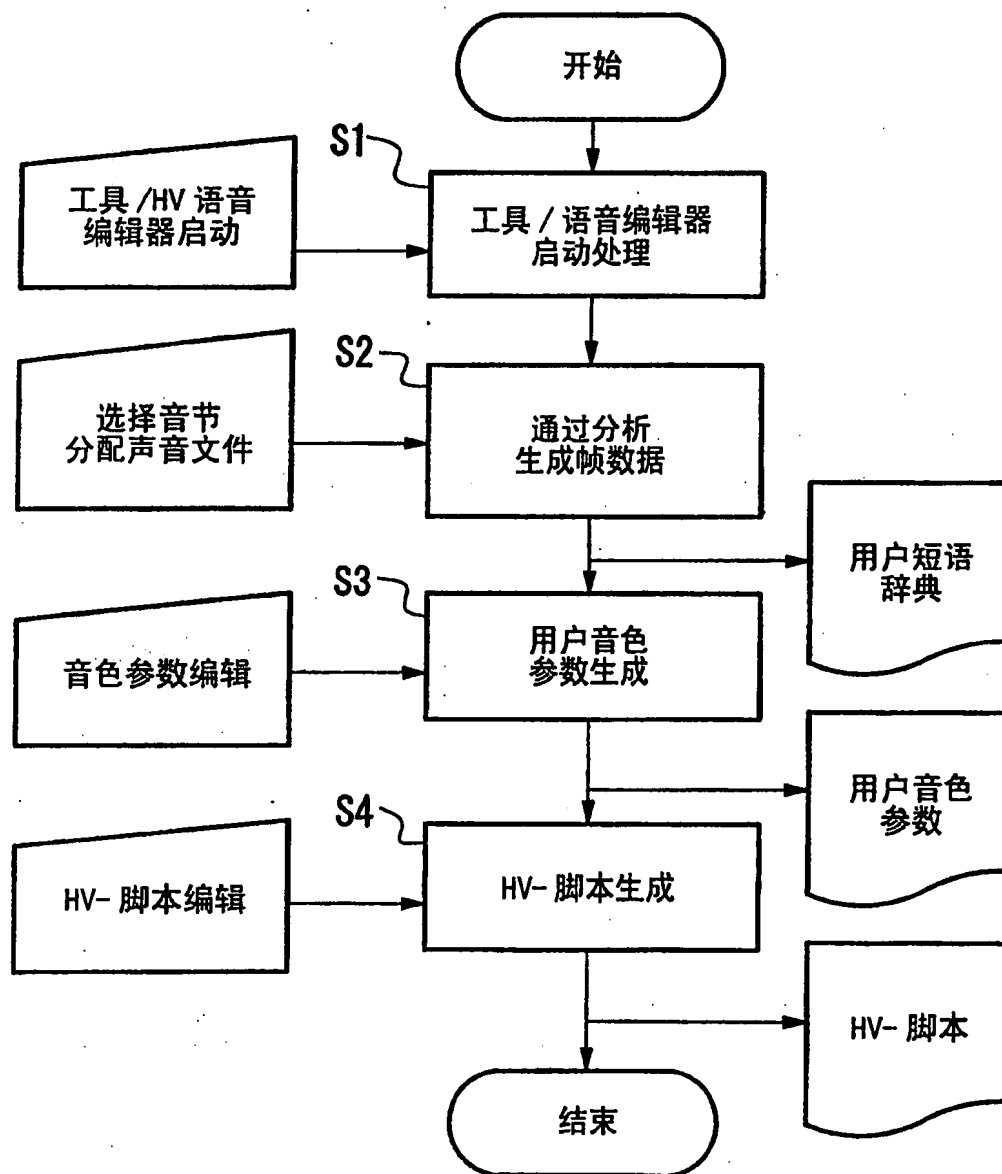


图 7

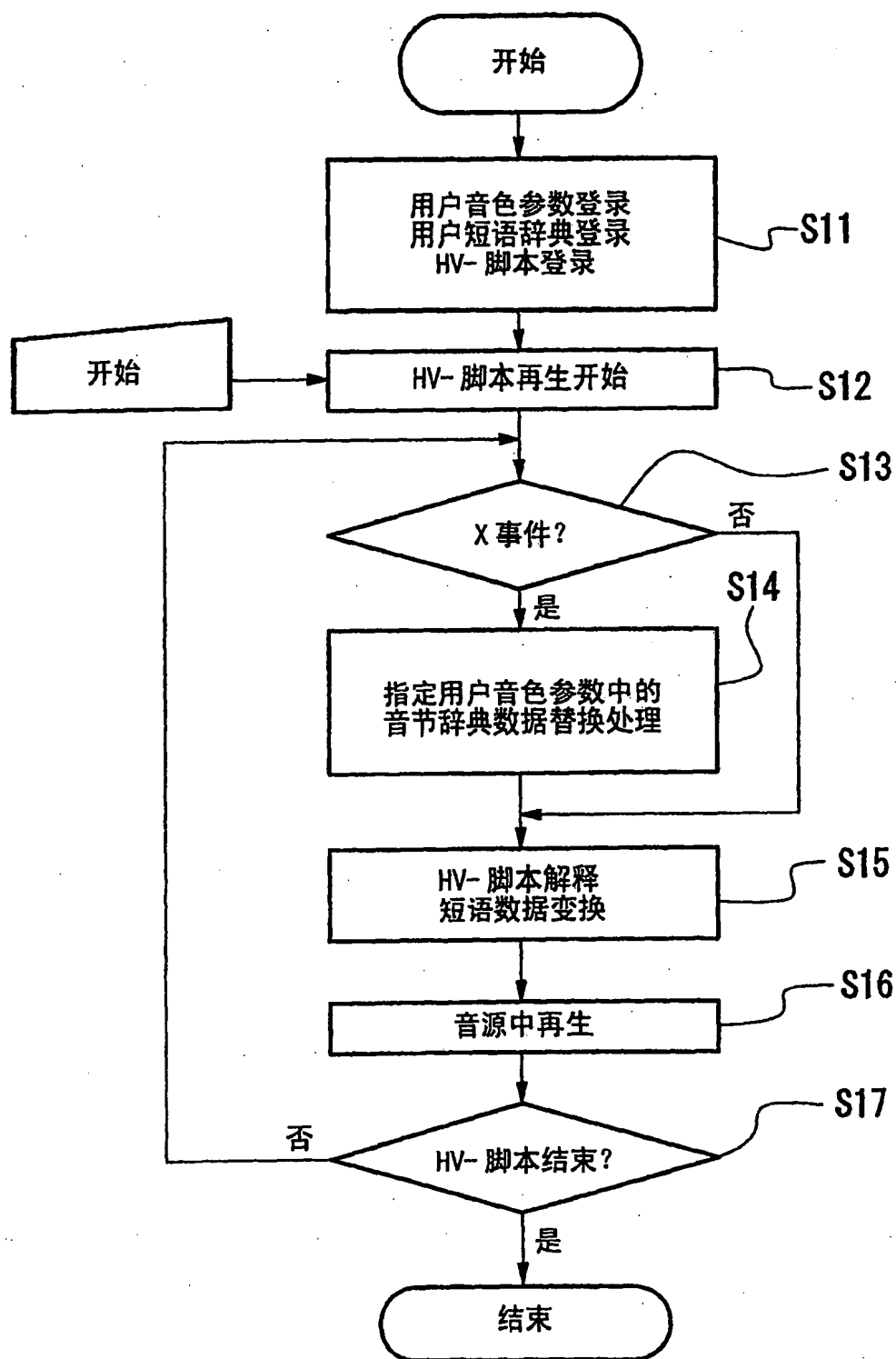


图 8

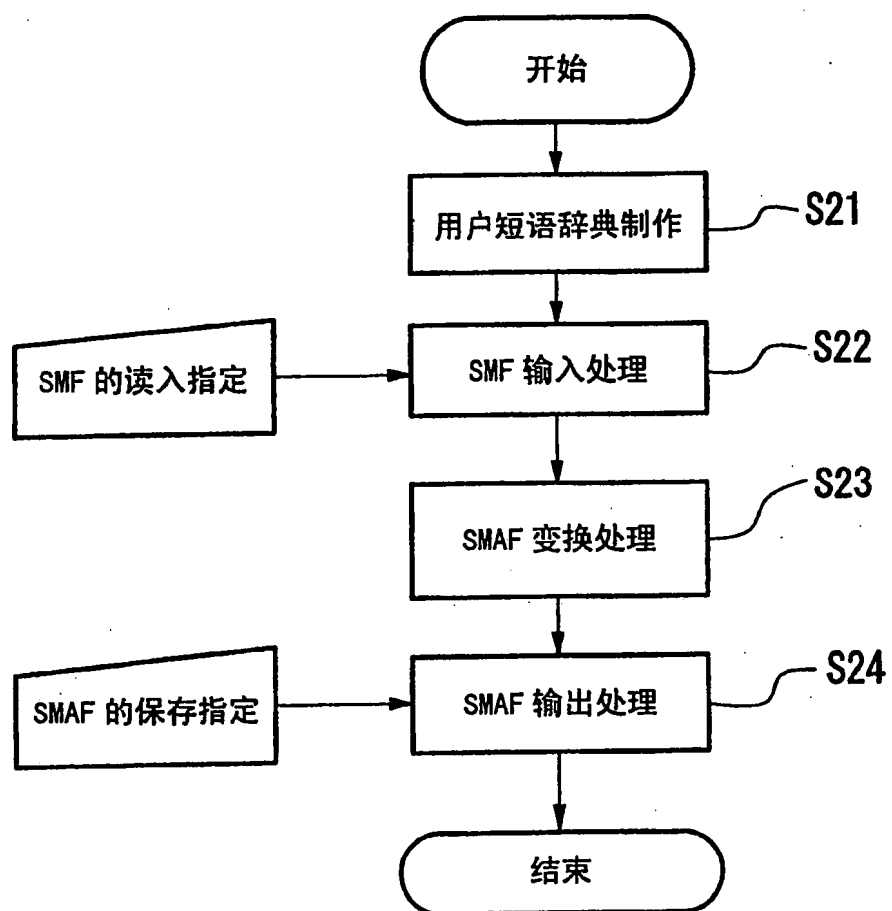


图 9

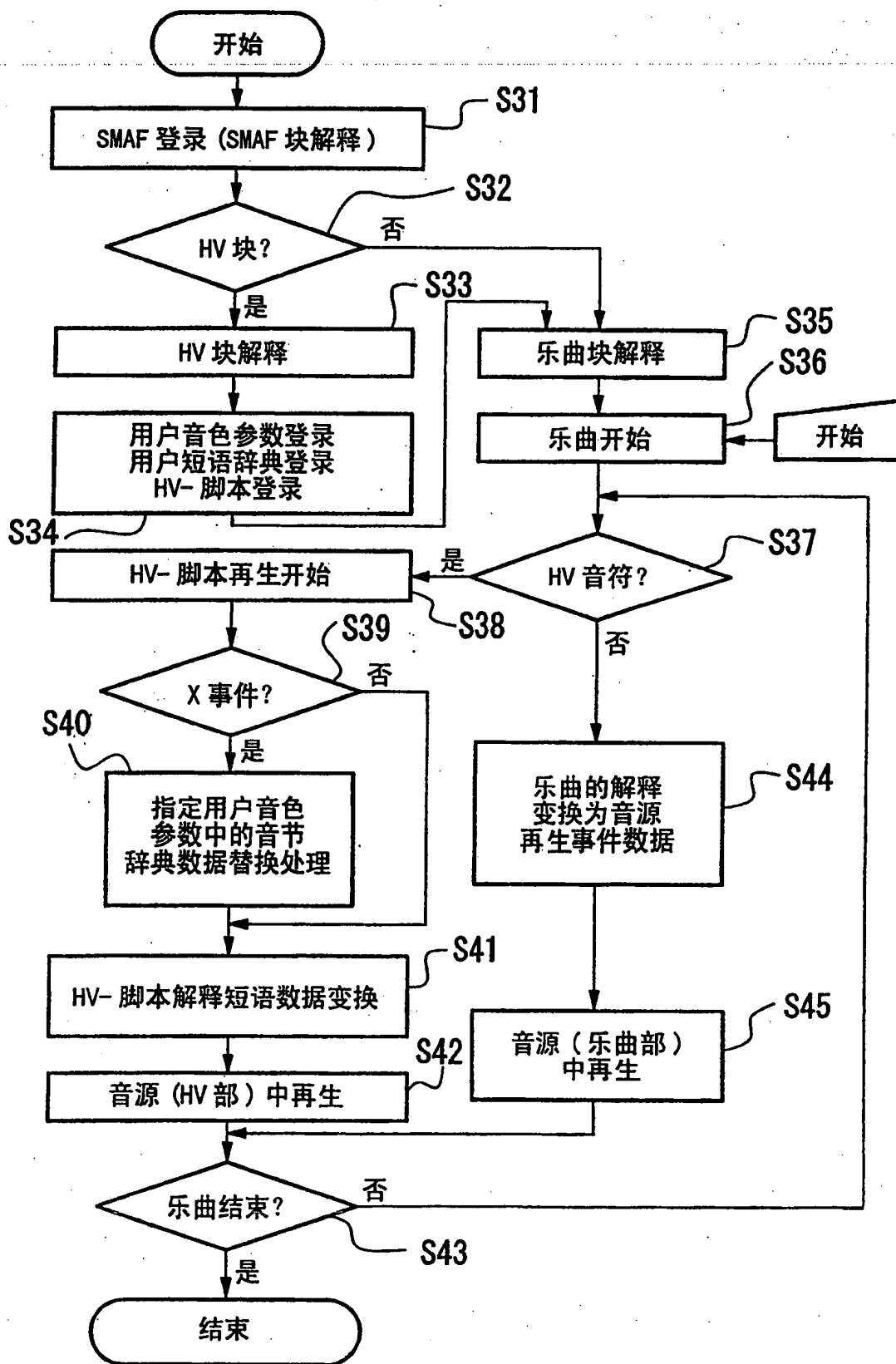


图 10

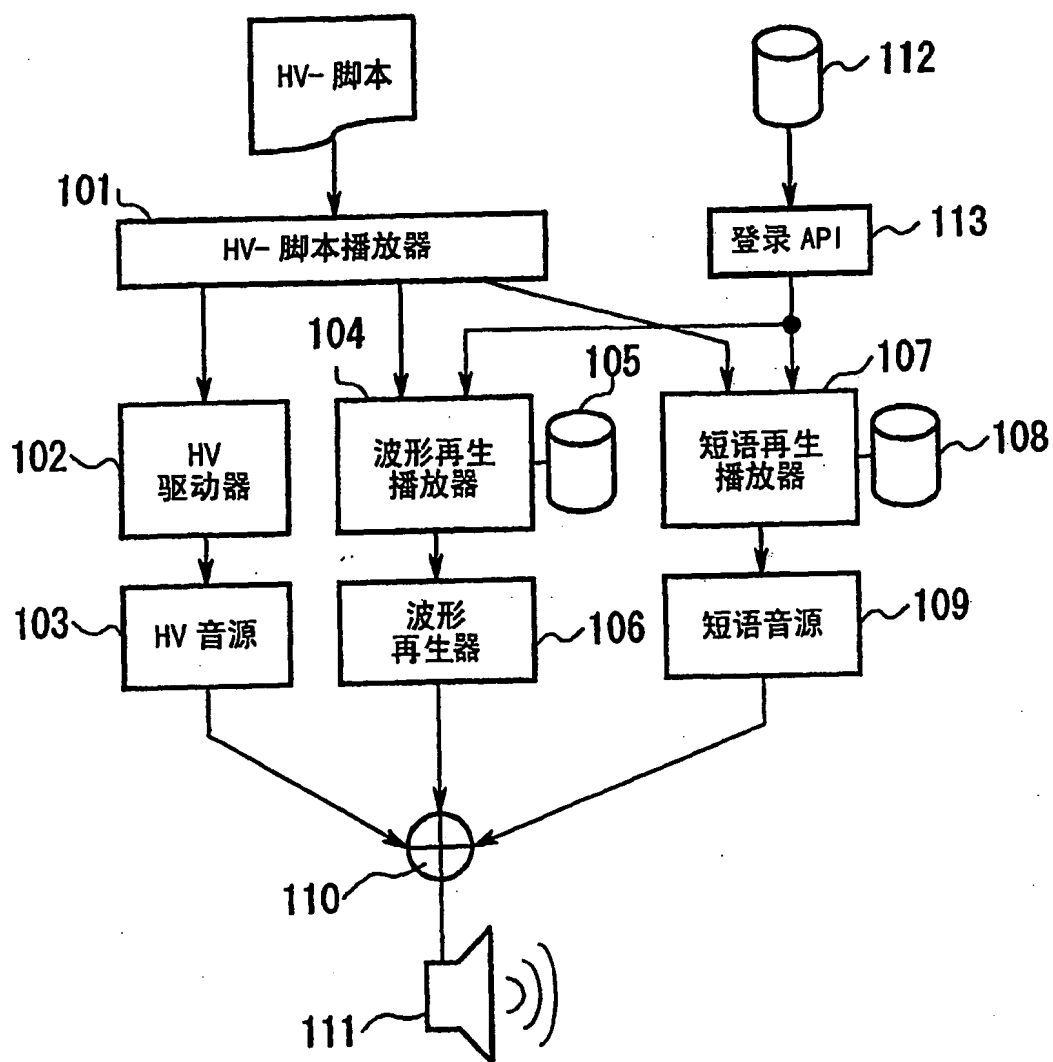


图 11

事件	类别	数据
D0	波形	缺省定义 0
⋮		
D63	短语	缺省定义 63
O0	波形	用户定义 0
⋮		
O63	短语	用户定义 63

图 12

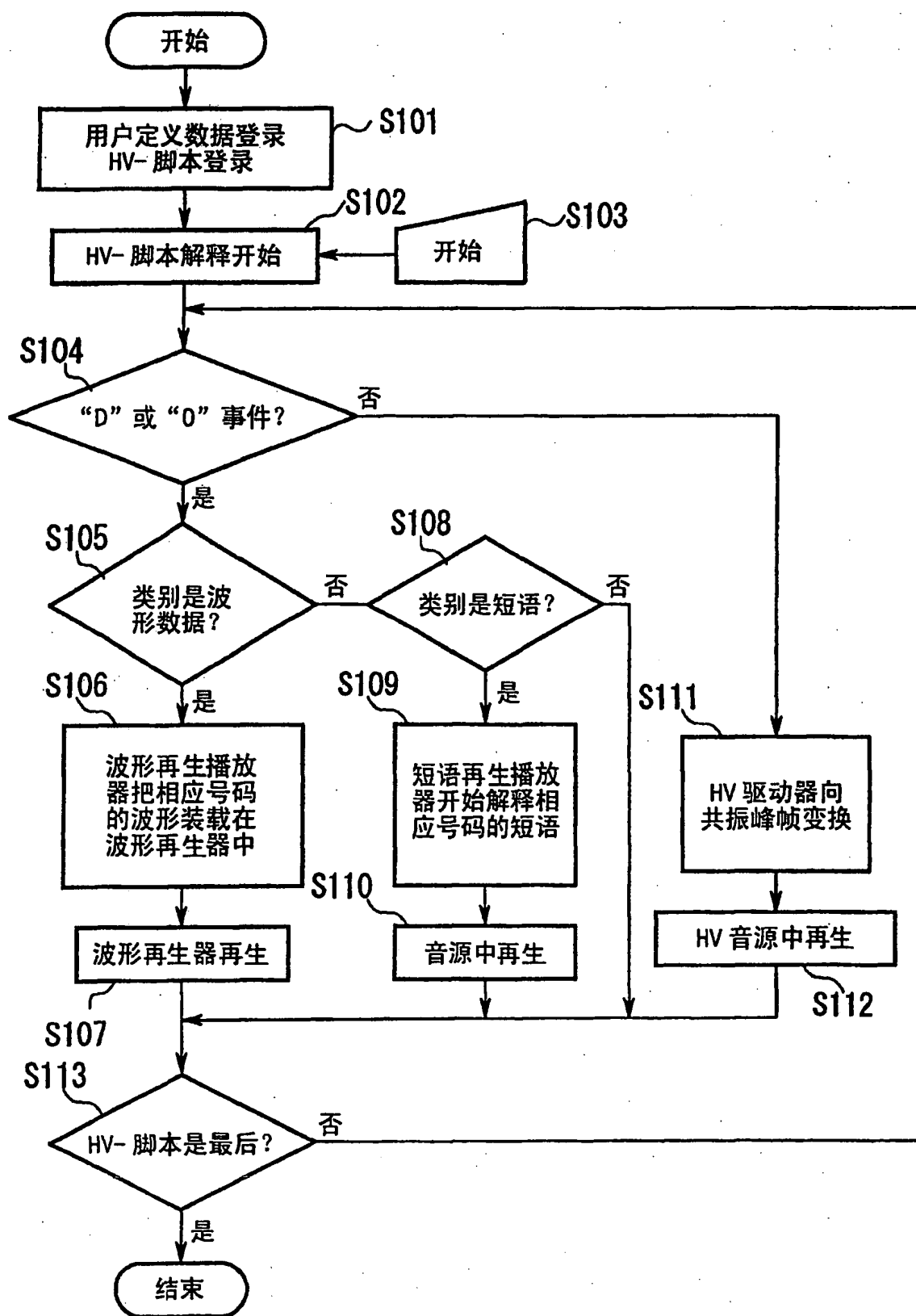


图 13

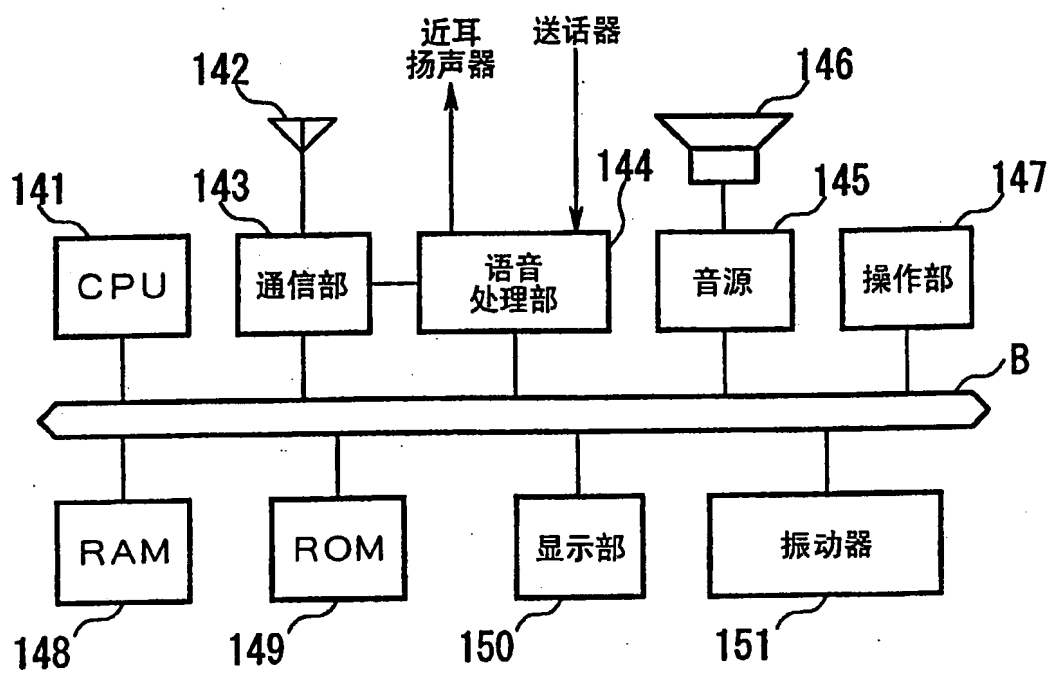


图 14

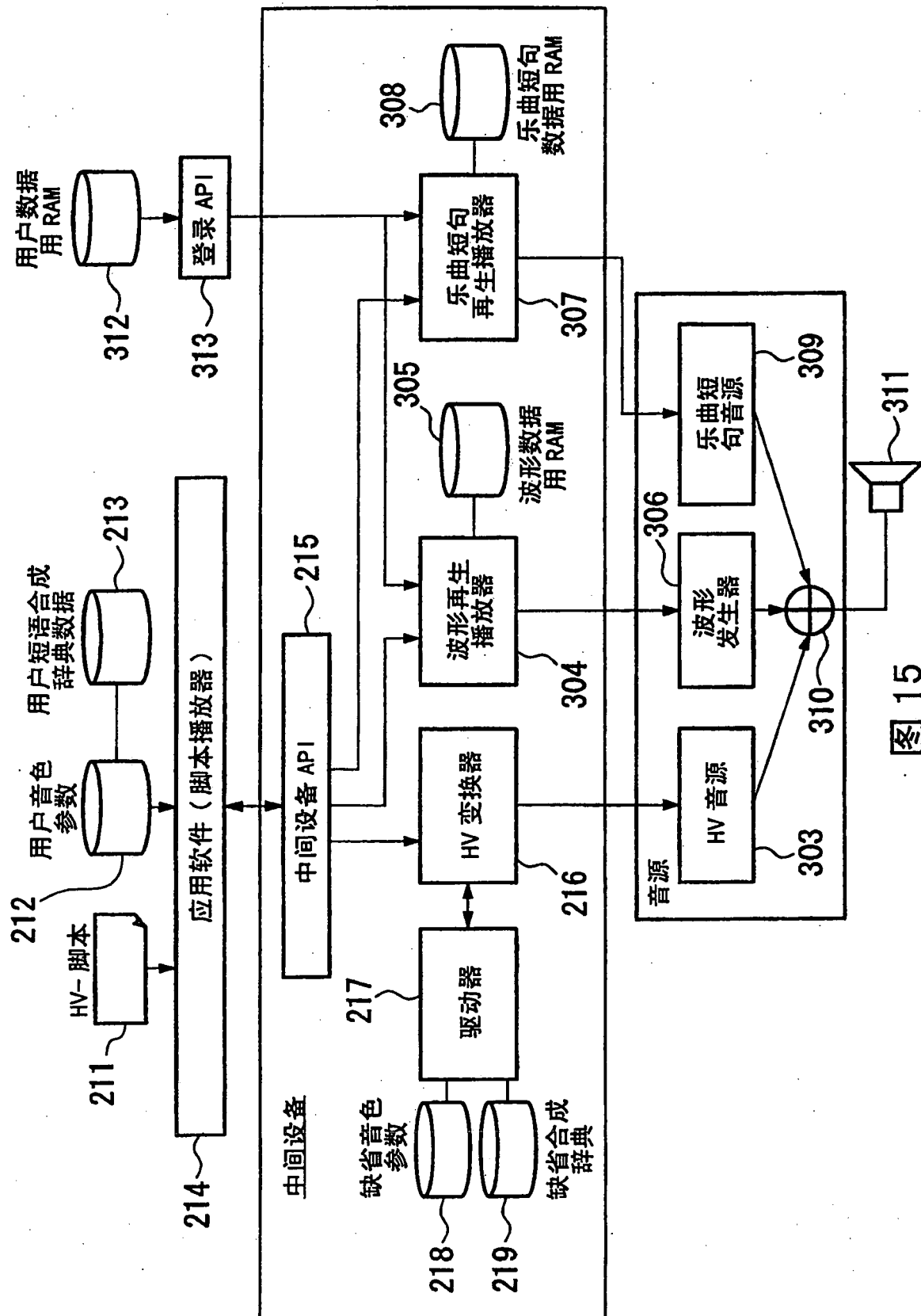


图 15

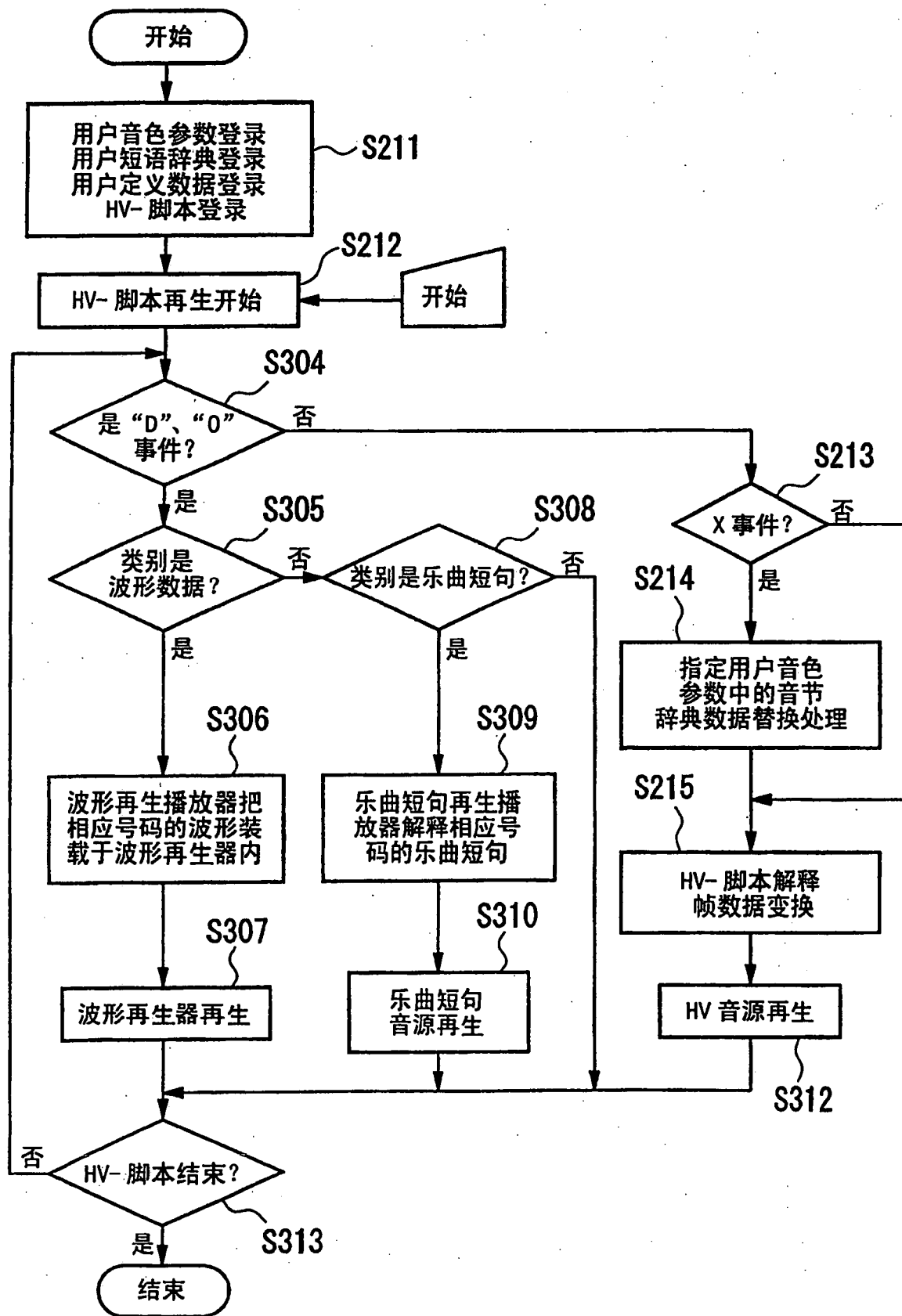


图 16