(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 1761993 B
(45)授权公告日 2010.05.05

(21)申请号 200480007616.6

(22)申请日 2004.03.19

(30)优先权数据
079152/2003 2003.03.20 JP

(85)PCT申请进入国家阶段日
2005.09.20

(86)PCT申请的申请数据
PCT/JP2004/003759 2004.03.19

(87)PCT申请的公布数据
WO2004/084175 JA 2004.09.30

(73)专利权人 索尼株式会社
地址 日本东京

(72)发明人 小林贤一郎

(74)专利代理机构 中国国际贸易促进委员会专利商标事务所 11038
代理人 党建华

(51)Int.Cl.
G10H 1/00(2006.01)
G10H 7/00(2006.01)
G10L 13/08(2006.01)
A63H 5/00(2006.01)

(56)对比文件
JP 20002-132281 ,2002.05.09, 全文 .
JP 10-319955 ,1998.12.04, 全文 .
JP 6-337690 ,1994.12.06, 全文 .
JP 2001-282269 ,2001.10.12, 全文 .
JP 2002-311952 ,2002.10.25, 全文 .

审查员 韦斌

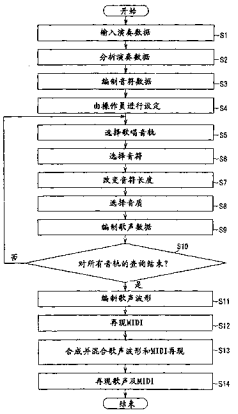
权利要求书 2 页 说明书 11 页 附图 9 页

(54)发明名称

歌声合成方法和设备以及机器人设备

(57)摘要

本发明涉及歌声合成方法和设备、程序、记录介质以及机器人设备，其中具体公开了一种利用演奏数据合成歌声的歌声合成方法。接收的演奏数据被分析为音调和音长以及歌词的音乐信息(S2,S3)。如果从被分析的音乐信息选择与歌词相应的音轨(S5)，从该音轨选择将被分配歌声的音符。



CN 1761993 B

1. 一种用于合成歌声的方法，包括：

分析步骤，所述分析步骤把演奏数据分析为音调和音长以及歌词的音乐信息；以及

歌声产生步骤，所述歌声产生步骤基于被分析的音乐信息而产生歌声；

所述歌声产生步骤基于包括在被分析音乐信息内的声音类型信息而决定所述歌声的类型。

2. 如权利要求 1 所述的歌声合成方法，其中，所述演奏数据是 MIDI 文件的演奏数据。

3. 如权利要求 2 所述的歌声合成方法，其中，所述歌声产生步骤基于包含在所述 MIDI 文件的演奏数据内的音轨中的乐器名或音轨名 / 序列名而决定歌声的类型。

4. 如权利要求 2 所述的歌声合成方法，其中，所述歌声产生步骤把从歌声的每个声音的音符开始时刻直到音符结束时刻的时间分配为歌声的一个声音，所述音符开始时刻是歌声的每个声音开始的时刻基准。

5. 如权利要求 4 所述的歌声合成方法，其中，利用是歌声的每个声音开始的时刻基准的所述 MIDI 文件的所述演奏数据中的音符开始时刻，在所述第一音符的音符结束之前有第二声音的音符开始作为叠加在所述第一音符上的音符的情况下，即使在所述第一声音的音符结束之前，所述歌声产生步骤使所述歌声的第一声音中断，所述歌声产生步骤也使歌声的所述第二声音在所述第二音符的音符开始时刻发音。

6. 如权利要求 5 所述的歌声合成方法，其中，如果在所述 MIDI 文件的所述演奏数据中在所述第一和第二音符之间没有重叠，所述歌声产生步骤就削减所述第一声音的音量，清楚地表现从歌声的第二声音开始的断点，在所述第一和第二音符之间有重叠并把所述第一和第二音符接合在一起以在音乐曲调上表现含糊发音的情况下，所述歌声产生步骤不削减音量。

7. 如权利要求 5 所述的歌声合成方法，其中，如果在所述第一和第二音符之间没有重叠，但在所述第一和第二音符之间只有比预定时间更短的声音中断间隔，所述歌声产生步骤就把所述第一声音的结束时刻移动到所述第二声音的开始时刻以把第一和第二声音接合在一起。

8. 如权利要求 4 所述的歌声合成方法，其中，如果在所述 MIDI 文件的演奏数据中有多个具有相同音符开始时刻的音符，所述歌声产生步骤就选择最高音调的音符作为歌声。

9. 如权利要求 4 所述的歌声合成方法，其中，如果在所述 MIDI 文件的演奏数据中有多个具有相同音符开始时刻的音符，所述歌声产生步骤就选择最低音调的音符作为歌声。

10. 如权利要求 4 所述的歌声合成方法，其中，如果在所述 MIDI 文件的演奏数据中有多个具有相同音符开始时刻的音符，所述歌声产生步骤就选择最大音量的音符作为歌声。

11. 如权利要求 4 所述的歌声合成方法，其中，如果在所述 MIDI 文件的演奏数据中有多个具有相同音符开始时刻的音符，所述歌声产生步骤就把这些音符处理成单独的声音部分，并向这些声音部分赋予相同的歌词，以产生不同音调值的歌声。

12. 如权利要求 4 所述的歌声合成方法，其中，如果从音符开始直到音符结束的时间长度比规定值更短，所述歌声产生步骤就不把该音符处理成歌唱主体。

13. 如权利要求 4 所述的歌声合成方法，其中，把从音符开始直到音符结束的时间长度扩展预定的比例，以产生歌声。

14. 如权利要求 13 所述的歌声合成方法，其中，以与乐器名相关联的形式设置用于改

变从音符开始直到音符结束的时间的所述预定比例的数据。

15. 如权利要求 4 所述的歌声合成方法，其中，所述歌声产生步骤在所述 MIDI 文件的所述演奏数据中在从音符开始直到音符结束的时间上增加预定时间，以产生歌声。

16. 如权利要求 15 所述的歌声合成方法，其中，以与乐器名相关联的形式设置用于改变从音符开始直到音符结束的时间的预定增加数据。

17. 如权利要求 4 所述的歌声合成方法，其中，所述歌声产生步骤改变从音符开始直到音符结束的时间，并且其中，由操作员设定用于改变所述时间的所述数据。

18. 如权利要求 2 所述的歌声合成方法，其中，所述歌声产生步骤从一个乐器名到下一乐器名地设定歌声类型。

19. 如权利要求 2 所述的歌声合成方法，其中，如果在所述 MIDI 文件的演奏数据中通过补丁而改变乐器的指定，即使在相同的音轨中，所述歌声产生步骤也改变歌声的类型。

20. 一种用于合成歌声的设备，包括：

分析装置，所述分析装置把演奏数据分析为音调和音长以及歌词的音乐信息；以及

歌声产生装置，所述歌声产生装置基于被分析的音乐信息而产生歌声；

所述歌声产生装置基于包括在被分析音乐信息内的声音类型信息而决定歌声的类型。

21. 如权利要求 20 所述的歌声合成设备，其中，所述演奏数据是 MIDI 文件的演奏数据。

22. 如权利要求 21 所述的歌声合成设备，其中，所述歌声产生装置基于包含在所述 MIDI 文件的演奏数据的音轨中的乐器名或音轨名 / 序列名而决定歌声的类型。

23. 如权利要求 21 所述的歌声合成设备，其中，所述歌声产生装置把从歌声的每个声音的音符开始时刻直到音符结束时刻的时间分配为歌声的一个声音，MIDI 文件的演奏数据中的所述音符开始时刻是歌声的每个声音开始的基准时刻。

24. 一种基于被提供的输入信息而执行动作的自主机器人设备，包括：

分析装置，所述分析装置把演奏数据分析为音调和音长以及歌词的音乐信息；以及

歌声产生装置，所述歌声产生装置基于被分析的音乐信息而产生歌声；

所述歌声产生装置基于包括在被分析音乐信息内的声音类型信息而决定所述歌声的类型。

25. 如权利要求 24 所述的用于合成歌声的机器人设备，其中，所述演奏数据是 MIDI 文件的演奏数据。

歌声合成方法和设备以及机器人设备

技术领域

[0001] 本发明涉及用于从演奏数据合成歌声的方法和设备、程序、记录介质以及机器人设备。

[0002] 本发明包含与2003年3月20日向日本专利办公室申请的日本专利申请JP-2003-079152有关的主题，此专利申请的全部内容在本文引作参考。

背景技术

[0003] 到目前为止知道例如通过计算机从给定歌唱数据合成歌声的技术。

[0004] 在相关技术领域中，MIDI（乐器数字接口）数据是被接受作为实际标准的代表性演奏数据。一般地，通过控制称作MIDI声源的数字声源而用MIDI数据产生音乐声音，其中，所述MIDI声源例如为由MIDI数据激发的声源，如计算机声源或电子乐器的声源。歌词数据可引入到MIDI文件，如SMF（标准MIDI文件），从而，可自动地编制具有歌词的音乐五线谱。

[0005] 例如，已经在日本专利特开平专利出版物H-11-95798中提出使用由歌声参数（特殊数据表示）或组成歌声的音位片段表现的MIDI数据的尝试。

[0006] 虽然这些相关技术试图用MIDI数据的数据形式来表现歌声，但是，此尝试仅仅是在控制乐器意义上的控制。

[0007] 而且，利用常规技术不纠正MIDI数据就把为乐器编制的MIDI数据译成歌曲是不可能的。

[0008] 另一方面，用于大声读电子邮件或主页的声音合成软件由包括本受让人在内的许多制造商销售。然而，读的方式是大声读文本的常规方式。

[0009] 使用电气或磁性操作来执行与包括人类的生命体相似的动作的机械设备称作机器人。机器人在日本的使用回溯到60年代末。当时使用的大多数机器人是工业机器人，如机械手或运输机器人，目的是使工厂的生产操作自动化或提供无人操作。

[0010] 近年来，正在进行应用机器人的开发，所述应用机器人适于支持人类生活，即在我们日常生活的各个方面支持人类活动，作为人类的伙伴。与工业机器人截然不同的是，应用机器人被赋予在我们日常生活的各个方面学习如何使它自己适合有个体差异的操作员或适应变化环境的能力。宠物型机器人或人形机器人正投入实际使用，其中，宠物型机器人模拟四足动物如狗或猫的身体机构或动作，人形机器人以人类用两条腿直立行走的身体机构或动作为模型进行设计。

[0011] 与工业机器人截然不同的是，应用机器人设备能执行以娱乐为中心的各种动作。为此，这些应用机器人设备有时称作娱乐机器人。在此类机器人设备中，有根据外部信息或内部状态而执行自主动作的机器人。

[0012] 用于自主机器人设备的人工智能(AI)是智力功能如推理或判断的人工实现。进一步试图人工实现诸如感觉或直觉的功能。在借助视觉装置或自然语言向外部表现人工智能的表现装置中，有借助声音的装置，作为使用自然语言的表现功能的实例。

[0013] 歌声的常规合成使用特殊类型的数据，或者即使使用 MIDI 数据，也不能有效地使用嵌入在其中的歌词数据，或者，不能唱为乐器准备的 MIDI 数据。

发明内容

[0014] 本发明的目的是提供一种有可能克服常规技术中固有问题的新型方法和设备。

[0015] 本发明的另一目的是提供一种合成歌声的方法和设备，从而，有可能通过利用演奏数据如 MIDI 数据而合成歌声。

[0016] 本发明的又一目的是提供一种合成歌声的方法和设备，其中，基于由 SMF 规定的 MIDI 数据的歌词信息而产生歌声，可自动地检验作为歌唱主体的声音串，从而在把声音串的音乐信息再现为歌声时，能实现‘含糊发音’或‘清晰发音’的音乐表现，并且其中，即使在不输入用于歌声的原始 MIDI 数据的情况下，也可从演奏数据选择作为歌唱主体的声音，并且，可调整声音长度或休止长度，以把音符或休止符转换为适于歌唱的音符或休止符。

[0017] 本发明的再一目的是提供一种使计算机执行歌声合成功能的程序和记录介质。

[0018] 根据本发明的歌声合成方法包括：分析步骤，所述分析步骤把演奏数据分析为音调和音长以及歌词的音乐信息；以及歌声产生步骤，所述歌声产生步骤基于被分析的音乐信息而产生歌声。歌声产生步骤基于包括在被分析音乐信息内的声音类型信息而决定歌声的类型。

[0019] 根据本发明的歌声合成设备包括：分析装置，所述分析装置把演奏数据分析为音调和音长以及歌词的音乐信息；以及歌声产生装置，所述歌声产生装置基于被分析的音乐信息而产生歌声。歌声产生装置基于包括在被分析音乐信息内的声音类型信息而决定歌声的类型。

[0020] 利用根据本发明的歌声合成方法和设备，有可能分析演奏数据，基于从音调、音长和声音速度或歌词得到的音符信息而产生与歌声有关的信息，从而产生歌声，其中，所述音调、音长和声音速度或歌词从被分析演奏数据获得，同时，基于与包含在被分析演奏数据内的声音类型有关的信息，有可能决定歌声的类型，从而允许以适合目标音乐曲调的音色和音质唱歌。

[0021] 根据本发明，演奏数据优选是 MIDI 文件如 SMF 的演奏数据。

[0022] 在此情况下，如果基于包括在 MIDI 文件的演奏数据的音轨中的乐器名或音轨名 / 序列名而决定歌声的类型，就可有利地利用 MIDI 数据。

[0023] 在向演奏数据的声音串分配歌词的成分时，例如，日本人希望把 MIDI 文件的演奏数据中从音符开始时刻直到音符结束时刻的时间间隔分配为歌声的一个声音，所述音符开始时刻是歌声的每个声音开始的基准。通过这样做，以演奏数据的每个音符一个歌声的速率唱出歌词，允许歌唱演奏数据的声音串。

[0024] 希望根据演奏数据的声音串中相邻音符的时间关系而调整歌声的声音互连的时间或方式。例如，如果第二音符的音符开始在时间上位于第一音符的音符结束之前，那么，甚至在第一音符的音符结束之前，就短暂地停止歌声第一声音的发音，并且，在第二声音的音符开始时刻发出第二声音，其中，第二音符是叠加在第一音符上的音符。如果在第一和第二音符之间没有重叠，就削减第一声音的音量，清楚地表现从第二声音开始的断点。如果在第一和第二音符之间有重叠，就把第一和第二音符接合在一起，而不削减第一声音的音量。

在前一种情况下，‘清晰地’唱歌，以相邻声音之间有间断地唱歌。在第二种情况下，‘含糊地’平滑地唱歌。如果在第一和第二音符之间没有重叠但在它们之间只有比预定时间间隔更短的声音中断时间间隔，第一声音的结束时刻就移到第二声音的开始时刻，在此时刻把第一和第二声音接合在一起。

[0025] 有在演奏数据中包括和音演奏数据的情况。例如，在 MIDI 数据的情况下，有在给定音轨或通道中记录和音演奏数据的情况。在存在此和音演奏数据的情况下，本发明考虑哪个声音串将作为歌词的主体。例如，如果在 MIDI 文件的演奏数据中有多个具有相同音符开始时刻的音符，就选择具有最高音调的音符作为歌唱主体的声音。这保证有利于歌唱所谓的女高音部分。可替换地，如果在 MIDI 文件的以上演奏数据中有多个具有相同音符开始时刻的音符，就选择具有最低音调的音符作为歌唱主体的声音。这保证歌唱所谓的低音部分。如果在 MIDI 文件的演奏数据中有多个具有相同音符开始时刻的音符，就选择具有最大指定音量的音符作为歌唱目标的声音。这保证歌唱主旋律或主题。还可替换地，如果在 MIDI 文件的以上演奏数据中有多个具有相同音符开始时刻的音符，各个音符就被处理成单独的声音部分，并向各个声音部分赋予相同的歌词，以产生不同音调值的歌声。这实现这些声音部分的合唱。

[0026] 还有在输入演奏数据中包括用于再现打击乐如木琴的乐声或短长度的改变声音的数据部分的情况。在此情况下，希望为歌唱调整歌声的长度。为此，如果在以上 MIDI 文件的演奏数据中从音符开始直到音符结束的时间比规定值更短，音符就不是歌唱的主体。或者，把以上 MIDI 文件的演奏数据中从音符开始直到音符结束的时间扩展预定的比例，以产生歌声。可替换地，在从音符开始直到音符结束的时间上增加预设时间，以产生歌声。希望以与乐器名相一致的形式设置和 / 或希望可由操作员设定用于改变从音符开始直到音符结束的时间的增加或比例的预设数据。

[0027] 优选地，按从一个乐器到另一乐器的歌声设定唱出的歌声类型。

[0028] 如果在 MIDI 文件的演奏数据中通过补丁而改变乐器的指定，即使在相同的音轨中，歌声设定步骤也希望在歌唱中途改变歌声的类型。

[0029] 根据本发明的程序允许计算机执行根据本发明的歌声合成功能。根据本发明的程序可由其中记录该程序的计算机读取。

[0030] 根据本发明的机器人设备是基于被提供的输入信息而执行动作的自主机器人设备，所述机器人设备包括 :分析装置，所述分析装置把演奏数据分析为音调和音长以及歌词的音乐信息 ;以及歌声产生装置，所述歌声产生装置基于被分析的音乐信息而产生歌声。歌声产生装置基于包括在被分析音乐信息内的声音类型信息而决定歌声的类型。这进一步提高作为娱乐机器人的机器人设备的性质。

附图说明

[0031] 图 1 为示出根据本发明的歌声合成设备的系统的框图。

[0032] 图 2 示出分析结果的音符信息的实例。

[0033] 图 3 示出歌声信息的实例。

[0034] 图 4 为示出歌声产生单元的结构的框图。

[0035] 图 5 示意性地示出用于解释歌声中音符长度调整的演奏数据中的第一和第二声

音。

[0036] 图 6 为示出根据本发明的歌声合成操作的流程图。

[0037] 图 7 为示出根据本发明的机器人设备的外观的透视图。

[0038] 图 8 示意性地示出机器人设备的自由度结构的模型。

[0039] 图 9 为示出机器人设备系统结构的框图。

具体实施方式

[0040] 参照附图详细解释本发明的优选实施例。

[0041] 图 1 示出根据本发明的歌声合成设备的示意性系统配置。应指出，预先假定本歌声合成设备例如用于机器人设备，其中，所述机器人设备至少包括感觉模型、语音合成装置和发音装置。然而，这不应解释为限制意义的，并且当然，本发明可应用于各种机器人设备以及除机器人之外的各种计算机 AI（人工智能）。

[0042] 在图 1 中，演奏数据分析单元 2 分析以 MIDI 数据为代表的演奏数据 1，分析输入的演奏数据，把该数据转换为音乐五线谱信息 4，所述音乐五线谱信息 4 表示包括在演奏数据中的音轨或通道的音调、音长和声音速度。

[0043] 图 2 示出转换为音乐五线谱信息的演奏数据（MIDI 数据）的实例。参照图 2，事件从一个音轨写到下一个音轨并从一个通道写到下一个通道。事件包括音符事件和控制事件。音符事件具有与产生时间（图 2 中的列‘时间’）、音调、长度和强度（速度）有关的信息。因而，音符串或声音串由音符事件序列定义。控制事件包括表示产生时间的数据、诸如颤音、演奏动态表现和控制内容的控制类型数据。例如，在颤音的情况下，控制内容包括表示声音脉动大小的‘深度’项、表示声音脉动周期的‘宽度’项、以及表示从声音脉动开始时刻（发声时刻）的‘延迟’项。用于特定音轨或通道的控制事件用于再现所述音轨或通道的音符串的音乐声，除非发生用于所述控制类型的新控制事件（控制变化）。而且，在 MIDI 文件的演奏数据中，可基于音轨而输入歌词。在图 2 中，在上半部表示的‘あるぅ日’（‘一天’，发‘a-ru-u-hi’音）是在音轨 1 中输入的歌词的一部分，而在下半部表示的‘あるぅ日’是在音轨 2 中输入的歌词的一部分。也就是说，在图 2 的实例中，歌词已经嵌入到被分析的音乐信息（音乐五线谱信息）中。

[0044] 在图 2 中，时间用“小节：拍：分段信号数量”表示，长度用“分段信号数量”表示，速度用数字‘0-127’表示，并且，音调用‘A4’代表 440Hz 而表示。另一方面，颤音的深度、宽度和延迟分别用数字‘0-64-127’表示。

[0045] 回到图 1，被转换的音乐五线谱信息 4 传递给歌词赋予单元 5。歌词赋予单元 5 根据音乐五线谱信息 4 而产生歌声信息 6，歌声信息 6 由用于声音的歌词以及与声音的长度、音调、速度和声调有关的信息组成，其中，所述声音的歌词与音符相匹配。

[0046] 图 3 示出歌声信息 6 的实例。在图 3 中，‘￥song ￥’为表示歌词信息开始的标签。标签‘￥PP，T10673075 ￥’表示 10673075μsec 的停顿，标签‘￥tdyna 110 649075 ￥’表示从前端开始 10673075μsec 的总速度，标签‘￥fine-100 ￥’表示细微的音调调整，与 MIDI 的微调相对应，并且，标签‘￥vibrato NRPN_dep ＝ 64 ￥’、‘￥vibrato NRPN_del ＝ 50 ￥’以及‘￥vibrato NRPN_rat ＝ 64 ￥’分别代表颤音的深度、延迟和宽度。标签‘￥dyna 100 ￥’代表不同声音的相对速度，并且，标签‘￥G4，T288461 ￥あ’代表具有

G4 音调和 288461μsec 长度的歌词元素‘あ’(发‘a’音)。图 3 的歌声信息从图 2 所示的音乐五线谱信息(MIDI 数据的分析结果)获得。图 3 的歌词信息从图 2 所示的音乐五线谱信息(MIDI 数据的分析结果)获得。从图 2 和 3 的比较可看出,用于控制乐器的演奏数据,如音乐五线谱信息,完全用于产生歌声信息。例如,对于歌词部分‘あるう日’中的组成元素‘あ’,其产生时间、长度、音调或速度包括在控制信息中或包括在音乐五线谱信息的音符事件信息中(参见图 2),并且与除‘あ’之外的其它歌唱属性一起直接使用,其中,所述歌唱属性例如为声音‘あ’的产生时间、长度、音调或速度,音乐五线谱信息中相同音轨或通道内的下一音符事件信息也直接用于下一歌词元素‘る’(发‘u’音),等等。

[0047] 参照图 1,歌声信息 6 传递给歌声产生单元 7,在此歌声产生单元 7 中,歌声产生单元 7 基于歌声信息 6 而产生歌声波形 8。从歌声信息 6 产生歌声波形 8 的歌声产生单元 7 例如按图 4 所示进行配置。

[0048] 在图 4 中,歌声节奏产生单元 7-1 把歌声信息 6 转换为歌声节奏数据。波形产生单元 7-2 通过基于音质的波形存储器 7-3 而把歌声节奏数据转换为歌声波形 8。

[0049] 作为具体实例,现在解释把歌词元素‘ら’(发‘ra’音)扩展为当前时间长度的情况。在不应用颤音情况下的歌声节奏数据可按下表 1 表示:

[0050] 表 1

[0051]

[标记]	[音调]	[音量]
0ra	050	066
1000aa		3960057
39600aa		4010048
40100aa		4060039
40600aa		4110030
41100aa		4160021
41600aa		4210012
42100aa		426003
42600aa		
43100a.		

[0052] 在上表中,[标记]代表各个声音(音位元素)的时间长度。也就是说,声音(音位元素)‘ra’具有从采样 0 到采样 1000 的 1000 个采样的时间长度,并且,初始声音‘aa’、下一声音‘ra’具有从采样 1000 到采样 39600 的 38600 个采样的时间长度。‘音调’代表以点音调表示的音调周期。也就是说,在采样点 0 的音调周期为 56 个采样。这里,不改变‘ら’的音调,从而,56 个采样的音调周期作用在全部采样上另一方面,‘音量’代表各个采样点每一个上的相对音量。也就是说,对于 100%的缺省值,在 0 采样点的音量为 66%,而在 39600 采样点的音量为 57%。在 40100 采样点的音量为 48%,在 42600 采样点的音量为 3%,等等。这实现‘ら’声音随着时间的衰减。

[0053] 另一方面,如果应用颤音,就编制下表 2 所示的歌声节奏数据:

[0054] 表 2

[0055]

[标记]	[音调]	[音量]
0ra	050	066
1000aa	100050	3960057
11000aa	200053	4010048
21000aa	400947	4060039
31000aa	600953	4110030
39600aa	801047	4160021
40100aa	1001053	4210012
40600aa	1201147	426003
41100aa	1401153	
41600aa	1602247	
42100aa	1802253	
42600aa	2003147	
43100a.	2203153	
	2404247	
	2604253	
	2804547	
	3004553	
	3205147	
	3405153	
	3606247	
	3806253	
	4007447	
	4207453	
	4301050	

[0056] 如上表的列‘音调’所示，在 0 采样点的音调周期和在 1000 采样点的音调周期都是 50 个采样，并且互相相等。在此时间间隔中，语音音调没有变化。从此时刻起，音调周期以大约 4000 个采样的周期（宽度）在 50±3 的范围内上下摆动，例如 :2000 采样点上 53 个采样的音调周期、4009 采样点上 47 个采样的音调周期以及 6009 采样点上 53 个采样的音调周期。以此方式，实现作为语音音调脉动的颤音。基于与歌声信息 6 中相应歌声元素如‘ら’有关的信息而产生列‘音调’的数据，所述信息具体为诸如 A4 的音调号、或诸如标签￥vibratoNRPN_dep ＝ 64 ￥’、‘￥vibrato NRPN_del ＝ 50 ￥’以及‘￥vibratoNRPN_rat ＝ 64 ￥’的颤音控制数据。

[0057] 基于以上歌声音位数据，波形产生单元 7-2 从基于音质的波形存储器 7-3 读出感兴趣音质的样本而产生歌声波形 8。在基于音质的波形存储器中已经储存不同音质的音位片段数据。当波形产生单元查询基于音质的波形存储器 7-3 时，波形产生单元 7-2 基于在歌声节奏数据中表示的音素序列、音调周期和音量而检索尽可能接近以上音素序列、音调周期和音量的音位片段数据。由此检索的数据被分片和排列，以产生语音波形数据。也就是说，音素数据例如以 CV（辅音 - 元音）、VCV 或 CVC 的形式按照不同的音质而储存在基于

音质的波形存储器 7-3 中。波形产生单元 7-2 基于歌声音位数据而按需要连接音素数据，并例如把适当的停顿、口音类型或语调附加到因此连接的数据上，以产生歌声波形 8。应指出，用于从歌声信息 6 产生歌声波形 8 的歌声产生单元不局限于歌声产生单元 7，并且，可以使用任何其它适当的歌声产生单元。

[0058] 回到图 1，演奏数据 1 传递给 MIDI 声源 9，MIDI 声源 9 接着基于演奏数据而产生音乐声。产生的音乐声是伴奏波形 10。

[0059] 歌声波形 8 和伴奏波形 10 传递给适于使两个波形互相合成和混合的混合单元 11。

[0060] 混合单元 11 使歌声波形 8 和伴奏波形 10 合成，并且，把两个波形叠加在一起，以产生并再现因此叠加的波形。因而，基于演奏数据 1，通过歌声及其附属的伴奏而再现音乐。

[0061] 歌词赋予单元 5 借助音轨选择器 12，基于音乐五线谱信息 4 中描述的音乐信息的任何音轨名 / 序列名、或乐器名而选择作为歌声主体的音轨。例如，如果声音或语音类型如‘女高音’被指定为音轨名，就直接确定该音轨是歌声的音轨。在诸如‘小提琴’的乐器的情况下，由操作员指定的音轨是歌声的主体。然而，如果操作员没有指定，情况就不是这样。在歌声主体数据 13 中包含给定音轨是否为歌声主体的信息，其内容可由操作员修改。

[0062] 另一方面，可由音质设定单元 16 设定哪一个音质应用于事先选择的音轨。在指定音质时，可从一个音轨到另一音轨并从一个乐器到另一个乐器地设定将要发音的声音类型。保留包括乐器名与音质之间相关性设定的信息，作为音质适应数据 19，并且，查询此音质适应数据，以选择例如与乐器名相关的音质。例如，作为歌声音质的音质‘女高音’、‘女低音 1’、‘女低音 2’、‘男高音 1’和‘男低音 1’分别与乐器名‘长笛’、‘单簧管’、‘中音萨克斯管’、‘低音萨克斯管’和‘巴松管’相关联。对于音质指定的优先次序，(a) 如果操作员已经指定音质，就应用因此指定的音质，并且 (b) 如果在音轨名 / 序列名中包含指定音质的字母 / 字符，就应用相关字母 / 字符串的音质。另外，(c) 在与乐器名有关的音质适应数据 19 中包含乐器名的情况下，就应用在音质适应数据 19 中描述的相应音质，并且，(d) 如果与以上条件不相关，就应用缺省音质。根据模式，可以或不可以应用此缺省音质。对于不应用缺省音质的模式，从 MIDI 再现乐器的声音。

[0063] 另一方面，如果在给定的 MIDI 音轨中乐器的指定已经通过修补而改变为控制数据，即使在相同的音轨中，也可根据音质适应数据 19 而中途改变歌声的音质。

[0064] 歌词赋予单元 5 基于音乐五线谱信息 4 而产生歌声信息 6。在此情况下，MIDI 数据中的音符开始时刻就用作歌曲的每个歌声开始的基准。从此时刻直到音符结束的声音被认为是一个声音。

[0065] 图 5 示出第一音符或第一声音 NT1 与第二音符或第二声音 NT2 之间的关系。在图 5 中，第一声音 NT1 的音符开始时刻表示为 t1a，第一声音 NT1 的音符结束时刻表示为 t1b，并且，第二声音 NT2 的音符开始时刻表示为 t2a。如上所述，歌词赋予单元 5 使用 MIDI 数据中的音符开始时刻作为歌曲中每个歌声的开始基准 (t1a 用作第一声音 NT1 的开始基准)，并且，把直到其音符结束为止的声音分配为一个歌声。这是歌词赋予的基础。因而，从一个声音到下一个声音地唱歌词，与 MIDI 数据的声音串中的每个音符的长度和音符开始时刻保持一致。

[0066] 然而，如果在第一声音 NT1 的音符开始和音符结束之间 (t1a ～ t1b) 之间有作为叠加声音的第二声音 NT2 的音符开始，即，如果 t1b ＞ t2a，音符长度改变单元 14 就改变歌

声的音符结束时刻，从而，歌声甚至在第一声音的音符结束之前就中断，并且，在第二声音NT2的音符开始时刻t2a发出下一歌声。

[0067] 如果在MIDI数据中在第一声音NT1和第二声音NT2之间没有重叠（t1a ＜ t2a），歌词赋予单元5就削减歌声中第一声音的音量，以便清楚地表现从歌声的第二声音开始的断点，以表现‘清晰发音’。如果相反在第一和第二声音之间有重叠，歌词赋予单元5就不削减音量，并把第一和第二声音接合在一起，在音乐曲调上表现‘含糊发音’。

[0068] 如果在MIDI数据中在第一声音NT1和第二声音NT2之间没有重叠，但只存在比储存于音符长度改变单元15中的预设时间更短的声音中断，音符长度改变单元14就把第一歌声的音符结束时刻移到第二歌声的音符开始时刻，以把第一和第二声音接合在一起。

[0069] 如果在MIDI数据中有多个其音符开始时刻相同（如t1a ＝ t2a）的音符或声音，歌词赋予单元5就使音符选择单元17根据音符选择模式18而从以下组中选择声音，作为歌声的主体，其中，所述组由具有最高音调的声音、具有最低音调的声音和具有最大音量的声音组成。

[0070] 在音符选择模式18中，可根据声音类型而设定将要选择具有最高音调的声音、具有最低音调的声音、具有最大音量的声音以及独立声音中的哪一个。

[0071] 如果在MIDI文件的演奏数据中有多个具有相同音符开始时刻的音符，并且在音符选择模式18中这些音符被设定为独立的声音，歌词赋予单元5就把这些声音处理为截然不同的声音部分，并且向这些声音赋予相同的歌词，以产生明显不同音调的歌声。

[0072] 如果从音符开始到音符结束的时间长度比通过音符长度改变单元14在音符长度改变数据15中设定的规定值更短，歌词赋予单元5就不使用该声音作为唱歌的主体。

[0073] 音符长度改变单元14通过在音符长度改变数据15中预设的比例，或通过增加规定时间而扩展从音符开始直到音符结束为止的时间。这些音符长度改变数据15以与乐器名匹配的形式保存在音乐五线谱信息中，并可由操作员设定。

[0074] 在前面已经结合歌词信息解释在演奏数据中包括歌词的情况。然而，本发明不局限于此配置。如果在演奏数据中不包括歌词，就可自动产生或由操作员输入可选歌词，如‘ら’或‘ぼん’（发‘bon’音），并且，通过音轨选择器或通过歌词赋予单元5选择作为歌词主体（音轨或通道）的演奏数据，以便歌词分配。

[0075] 图6示出歌声合成设备的总体操作的流程图。

[0076] 首先，输入MIDI文件的演奏数据1（步骤S1）。接着分析演奏数据1，并接着输入音乐五线谱数据4（步骤S2和S3）。随后向执行设定处理的操作员进行询问（步骤S4），其中，所述设定处理例如设定作为歌声主体的数据、选择音符的模式、改变音符长度的数据或用于处理音质的数据。在操作员还未执行设定的情况下，在后续处理中应用缺省设定。

[0077] 随后的步骤S5-S10表示用于产生歌声信息的循环。首先，通过音轨选择单元12选择作为歌词主体的音轨（步骤S5）。通过音符选择单元17从作为歌词主体的音轨确定将根据音符选择模式而分配给歌声的音符（步骤S6）。如果需要，通过音符长度改变单元14根据以上定义的条件而改变分配给歌声的音符的长度，如发音时刻或时间长度（步骤S7）。接着，通过歌词赋予单元5，基于在步骤S5-S8中获得的数据而准备歌声信息6（步骤S9）。

[0078] 接着，检查对所有音轨的查询是否已经结束（步骤S10）。如果查询还未结束，处理就返回到步骤S5，并且，如果查询已经结束，歌声信息6就传递给歌声产生单元7，以编制歌

声波形（步骤 S11）。

[0079] 接着，通过 MIDI 声源 9 再现 MIDI，以编制伴奏波形 10（步骤 S12）。

[0080] 通过到目前为止执行的处理，编制歌声波形 8 和伴奏波形 10。

[0081] 当两个波形互相合成时，混合单元 11 把歌声波形 8 和伴奏波形 10 叠加在一起，以形成被再现的输出波形 3（步骤 S13 和 S14）。此输出波形 3 通过未示出的声音系统输出，作为声信号。

[0082] 上述歌声合成功能例如包括在机器人设备中。

[0083] 以本发明实施例示出的用两条腿行走类型的机器人设备是在我们日常生活各个方面，如在我们的生活环境中，支持人类活动的应用机器人，并且能根据内部状态如愤怒、悲伤、快乐或幸福而动作。同时，这是能表现人类基本行为的娱乐机器人。

[0084] 参照图 7，机器人设备 60 由躯干单元 62 形成，躯干单元 62 在预定位置连接到头部单元 63、左右臂单元 64R/L 以及左右腿单元 65R/L，其中，R 和 L 分别代表表示右和左的后缀，以下相同。

[0085] 在图 8 中示意性地示出为机器人设备 60 设置的关节的自由度结构。支撑头部单元 63 的颈关节包括三个自由度，即颈关节偏转轴 101、颈关节俯仰轴 102 和颈关节翻滚轴 103。

[0086] 组成上肢的臂单元 64R/L 由肩关节俯仰轴 107、肩关节翻滚轴 108、上臂偏转轴 109、肘关节俯仰轴 110、前臂偏转轴 111、腕关节俯仰轴 112、腕关节翻滚轴 113 和手单元 114 组成。手单元 114 实际上是包括多个手指的多关节多自由度结构。然而，由于手单元 114 的动作作用于或者影响机器人设备 60 的姿势控制或行走控制，因此，在本文描述中假设手单元具有零自由度。结果，每个臂单元都设置七个自由度。

[0087] 躯干单元 62 也具有三个自由度，即，躯干俯仰轴 104、躯干翻滚轴 105 和躯干偏转轴 106。

[0088] 形成下肢的每个腿单元 65R/L 都由臀关节偏转轴 115、臀关节俯仰轴 116、臀关节翻滚轴 117、膝关节俯仰轴 118、踝关节俯仰轴 119、踝关节翻滚轴 120、以及腿单元 121 组成。在本文描述中，臀关节俯仰轴 116 和臀关节翻滚轴 117 的交叉点规定机器人设备 60 的臀关节位置。尽管实际上人类的腿单元 121 是包括脚底的结构，其中，脚底具有多个关节和多个自由度，但是，假设机器人设备的脚底是零自由度的。结果，每条腿具有六个自由度。

[0089] 总之，机器人设备 60 全部具有总计 3+7×2+3+6×2 ＝ 32 个自由度。然而，应指出，娱乐机器人设备的自由度的数量不局限于 32，从而，可根据设计或制造中的约束条件或根据要求的设计参数而适当地增加或减少自由度的数量，即，关节数量。

[0090] 实际上使用执行器来安装上述机器人设备 60 拥有的上述自由度。考虑到消除外观上过度的肿胀以接近人体自然形状的要求、以及对因两条腿行走导致的不稳定结构进行姿势控制的要求，希望执行器尺寸小且重量轻。更优选执行器设计和构造为直接传动耦合类型的小尺寸 AC 伺服执行器，其中，伺服控制系统布置为一个芯片并安装在电动机单元中。

[0091] 图 9 示意性地示出机器人设备 60 的控制系统结构。参照图 9，控制系统由思维控制模块 200 以及动作控制模块 300 组成，其中，思维控制模块 200 根据用户输入而动态地负责情绪判断或感觉表达，动作控制模块 300 控制机器人设备 60 全部躯体的协同动作，如驱

动执行器 350。

[0092] 思维控制模块 200 是独立驱动的信息处理设备，它由执行计算与情绪判断或感觉表达的 CPU（中央处理单元）211、RAM（随机存取存储器）212、ROM（只读存储器）213、以及外部存储装置（如硬盘驱动器）214 组成，并且能在模块内执行自主式处理。

[0093] 此思维控制模块 200 根据外部的刺激，如从图像输入装置 251 输入的图像数据或从声音输入装置 252 输入的声音数据，而决定机器人设备 60 当前的感觉或意向。图像输入装置 251 例如包括多个 CCD（电荷耦合装置）照相机，而声音输入装置 252 包括多个麦克风。

[0094] 思维控制模块 200 基于决定而发出对动作控制模块 300 的命令，以便执行动作的行为序列，即四肢的动作。

[0095] 动作控制模块 300 是独立驱动的信息处理设备，它由控制机器人设备 60 全部躯体的协同动作的 CPU（中央处理单元）311、RAM 312、ROM 313、以及外部存储装置（如硬盘驱动器）314 组成，并且能在模块内执行自主式处理。外部存储装置 314 能储存动作表，包括脱机计算的行走方案以及目标 ZMP 轨迹。应指出，ZMP 是在地板表面上在行走过程中从地板作用的反作用力的力矩等于零的点，而 ZMP 轨迹是在机器人设备 60 的行走周期中 ZMP 移动的轨迹。对于 ZMP 的概念以及应用 ZMP 作为行走机器人稳定程度的检验标准，参照 Miomir Vukobratovic 的“有腿移动机器人 (Legged LocomotionRobots)”，以及 Ichiro KATO 等的“行走机器人和人造腿 (WalkingRobot and Artificial Legs)”，NIKKAN KOGYO SHIMBUN-SHA 出版。

[0096] 通过总线接口 (I/F)301 连接到动作控制模块 300 的例如有执行器 350、姿势传感器 351、地板接触确认传感器 352、353、以及电源控制装置 354，其中，执行器 350 分布在图 9 所示机器人设备 60 的全部躯体上，用于实现自由度；姿势传感器 351 用于测量躯干单元 62 的倾斜姿势；地板接触确认传感器 352、353 用于检测左右脚的脚底的飞跃状态或站立状态；电源控制装置 354 用于监督诸如电池的电源。例如通过组合加速传感器和陀螺仪传感器而形成姿势传感器 351，同时，地板接触确认传感器 352、353 中的每一个都由近程传感器或微型开关形成。

[0097] 思维控制模块 200 和动作控制模块 300 在公共平台上形成，并且通过总线接口 201、301 互连。

[0098] 动作控制模块 300 控制由各个执行器 350 产生的全部躯体的协同动作，用于实现由思维控制模块 200 命令的行为。也就是说，CPU 311 从外部存储装置 314 中提取出与思维控制模块 200 所命令行为一致的行为方案，或者在内部产生该行为方案。CPU 311 根据指定的动作方案而设定脚 / 腿动作、ZMP 轨迹、躯干动作、上肢动作、水平位置和腰部高度，同时向各个执行器发送命令值，以命令执行与设定内容一致的动作。

[0099] CPU 311 还基于姿势传感器 351 的控制信号而检测机器人设备 60 的躯干单元 62 的姿势或倾斜，同时，通过地板接触确认传感器 352、353 的输出信号检测腿单元 65R/L 是处于飞跃状态还是处于站立状态，以便适应性地控制机器人设备 60 全部躯体的协同动作。

[0100] CPU 311 还控制机器人设备 60 的姿势或动作，从而，ZMP 位置总是指向 ZMP 稳定区的中心。

[0101] 动作控制模块 300 适于向思维控制模块 200 返回已经实现与思维控制模块 200 所做决定保持一致的行为的程度，即处理状态。

[0102] 以此方式，机器人设备 60 能基于控制程序而核实自己的状态和周围的状态，以执行自主行为。

[0103] 在此机器人设备 60 中，例如在思维控制模块 200 的 ROM 213 中驻留已经实施上述歌声合成功能的程序，包括数据。在此情况下，用于合成歌声的程序由思维控制模块 200 的 CPU 211 执行。

[0104] 通过向机器人设备提供上述歌声合成功能，新获得机器人设备对着伴奏唱歌的表现能力，结果是该机器人设备作为娱乐机器人的性质得到增强，进一步密切机器人设备与人类的关系。

[0105] 工业应用

[0106] 对于根据本发明的歌声合成方法和设备，其中，演奏数据被分析为音调和音长的音乐信息以及歌词的音乐信息，基于被分析的音乐信息而产生歌声，并且，其中，基于包含在被分析音乐信息内的声音类型信息而确定歌声的类型，有可能分析给定的演奏数据，以根据音符信息而产生歌声信息，以便根据歌声信息而产生歌声，其中，所述音符信息是基于从分析得到的歌词或音调、音长或声音速度。还有可能基于与包含在被分析音乐信息内的声音类型有关的信息而确定歌声类型，从而，有可能以适合感兴趣音乐曲调的音色和音质唱歌。结果，不必增加在到目前为止只通过乐器声音而编制或表现音乐时的任何特殊信息而再现歌声，因此，可较大地提高音乐表现力。

[0107] 根据本发明的程序允许计算机执行本发明的歌声合成功能。在根据本发明的记录介质上记录此程序，并且，此介质是计算机可读的。

[0108] 对于根据本发明的程序和记录介质，其中，演奏数据被分析为音调和音长的音乐信息以及歌词的音乐信息，基于被分析的音乐信息而产生歌声，并且，其中，基于包含在被分析音乐信息内的声音类型信息而确定歌声的类型，可分析演奏数据，基于音符信息而产生歌声信息，并且，基于因此产生的歌声信息而产生歌声，其中，所述音符信息是基于从分析得到的音调、音长或声音速度和歌词。而且，通过基于与包含在被分析音乐信息内的声音类型有关的信息而决定歌声类型，以适合目标音乐曲调的音色和音质唱歌。

[0109] 根据本发明的机器人设备能实现根据本发明的歌声合成功能。也就是说，对于根据本发明的基于被提供的输入信息而执行动作的自主机器人设备，演奏数据被分析为音调和音长的音乐信息以及歌词的音乐信息，基于被分析的音乐信息而产生歌声，并且，其中，基于包含在被分析音乐信息内的声音类型信息而确定歌声的类型，可分析演奏数据，基于音符信息而产生歌声信息，并且，基于因此产生的歌声信息而产生歌声，其中，所述音符信息是基于从分析得到的音调、音长和声音速度以及歌词。而且，通过基于与包含在被分析音乐信息内的声音类型有关的信息而决定歌声类型，以适合目标音乐接合的音色和音质唱歌。结果是可提高机器人设备的表现力，作为娱乐机器人的机器人设备的性质得到增强，进一步密切机器人设备与人类的关系。

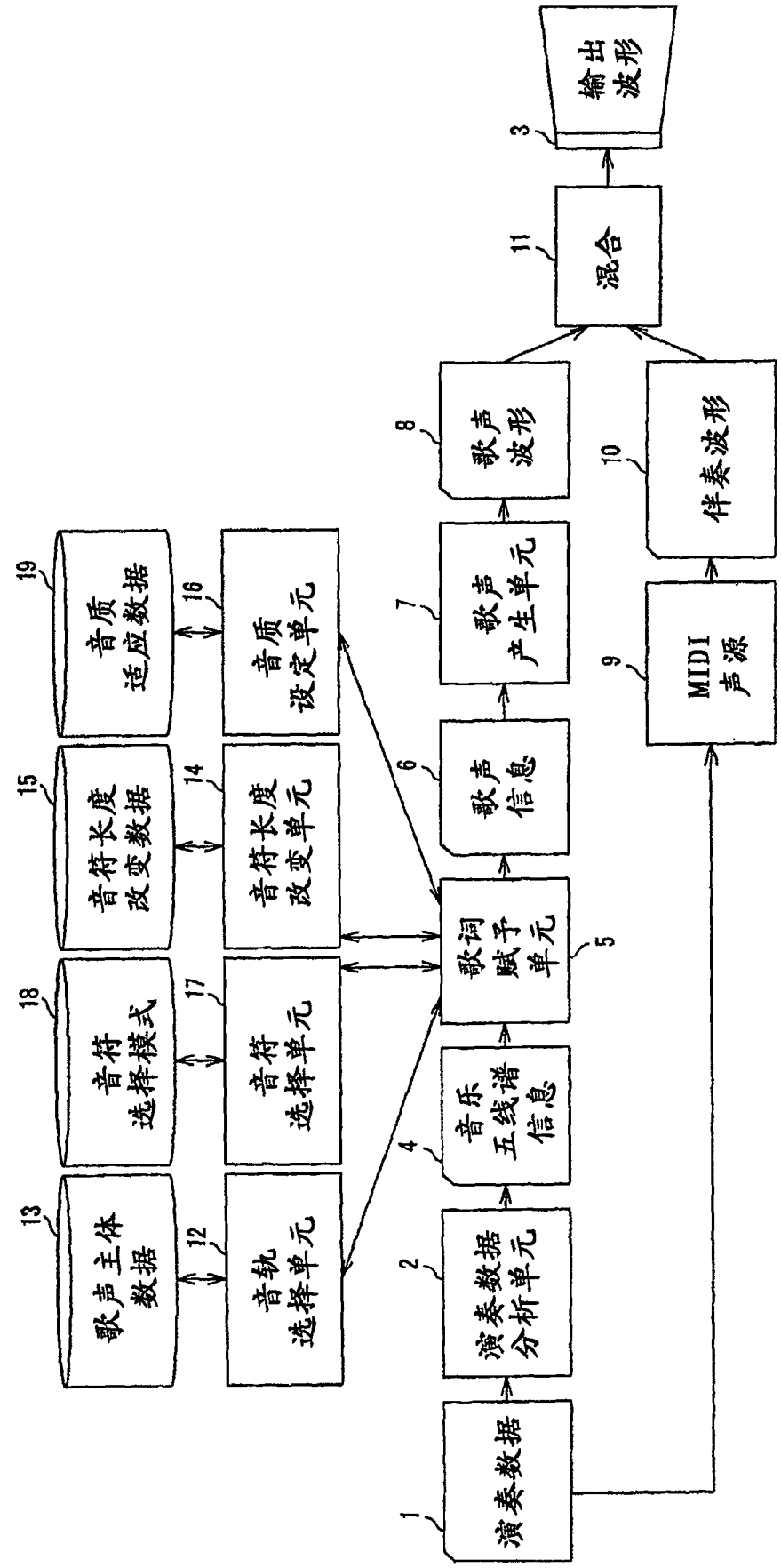
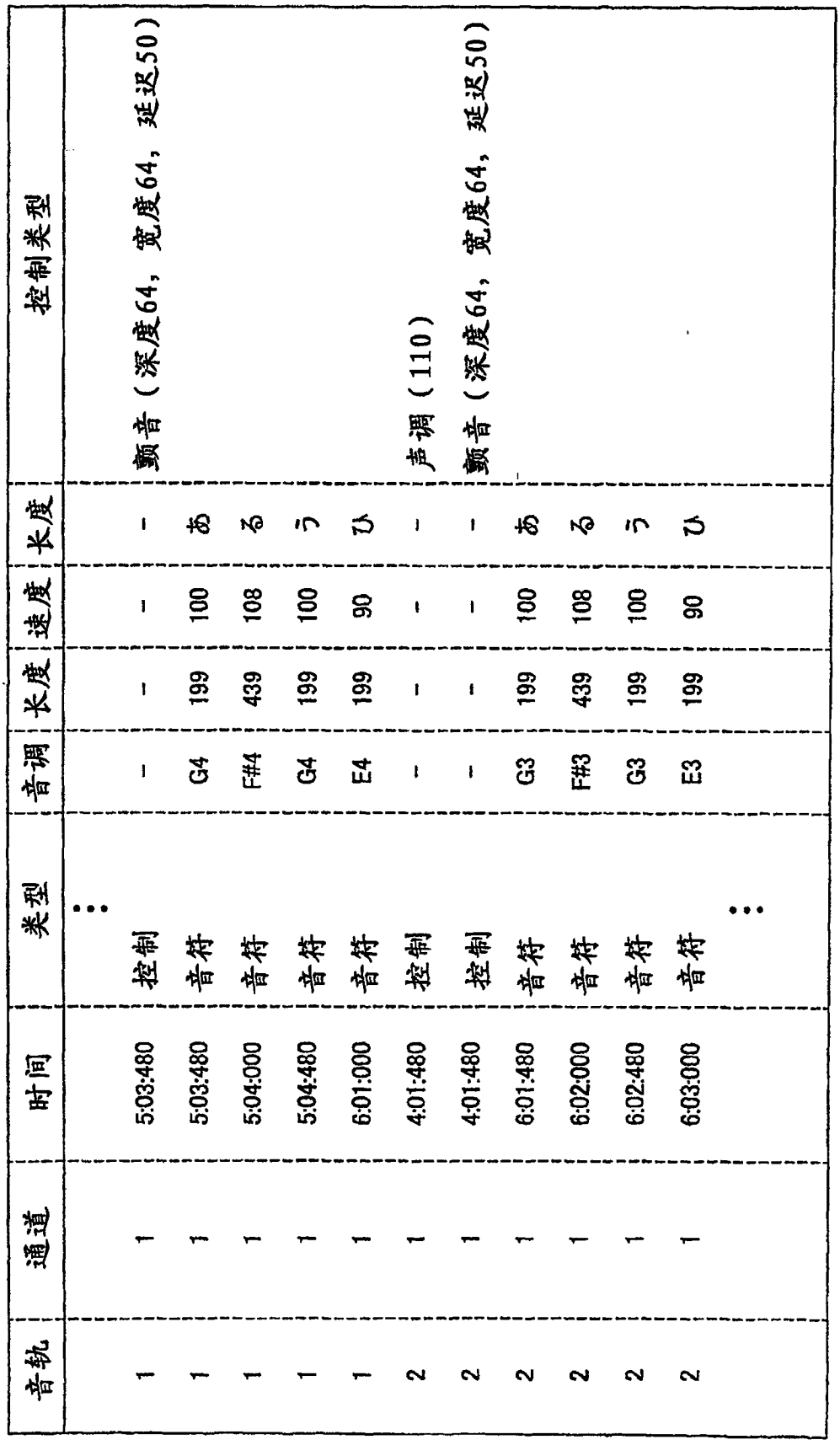

演奏数据
1
演奏数据分析单元
2
音乐五线谱信息
4
歌词赋予单元
5
歌声信息
6
歌声产生单元
7
歌声波形
8
MIDI声源
9
伴奏波形
10
混合
11
输出波形
3
音轨选择单元
12
歌声主体数据
13
音符选择单元
17
音符选择模式
18
音符长度改变单元
14
音符长度改变数据
15
音质设定单元
16
音质适应数据
19

图 1



音轨	通道	时间	类型	音调	长度	速度	长度	控制类型
			...					
1	1	5:03:480	控制	–	–	–	–	颤音（深度64，宽度64，延迟50）
1	1	5:03:480	音符	G4	199	100	あ	
1	1	5:04:000	音符	F#4	439	108	る	
1	1	5:04:480	音符	G4	199	100	う	
1	1	6:01:000	音符	E4	199	90	ひ	
2	1	4:01:480	控制	–	–	–	–	声调（110）
2	1	4:01:480	控制	–	–	–	–	颤音（深度64，宽度64，延迟50）
2	1	6:01:480	音符	G3	199	100	あ	
2	1	6:02:000	音符	F#3	439	108	る	
2	1	6:02:480	音符	G3	199	100	う	
2	1	6:03:000	音符	E3	199	90	ひ	
			...					

图 2

¥song¥ ←—— 歌声数据的开始
¥PP,T10673075¥ ←—— 10673075μsec 停顿
¥tdyna 110 649075¥ ←—— 从头开始 10673075μsec 的总速度
¥fine-100¥ ←—— 细微的音调调整（与 MIDI 的微调相同）
¥vibrato NRPN_dep=64¥ ←—— 颤音
¥vibrato NRPN_del=50¥
¥vibrato NRPN_rat=64¥
¥dyna 100¥ ←—— 每个声音的强/弱
¥G4,T288461¥あ ←—— 具有 G4 音调和 288461μsec 音长的声音，歌词为'あ'
¥dyna 108¥
¥Gb4,T288462¥る
¥dyna 100¥
¥G4,T288461¥う
¥dyna 90¥
¥E4,T219592¥ひ
¥PP,T1222716¥
¥dyna 100¥
¥E4,T144231¥も
¥dyna 98¥
¥E4,T144230¥り
⋮

图 3

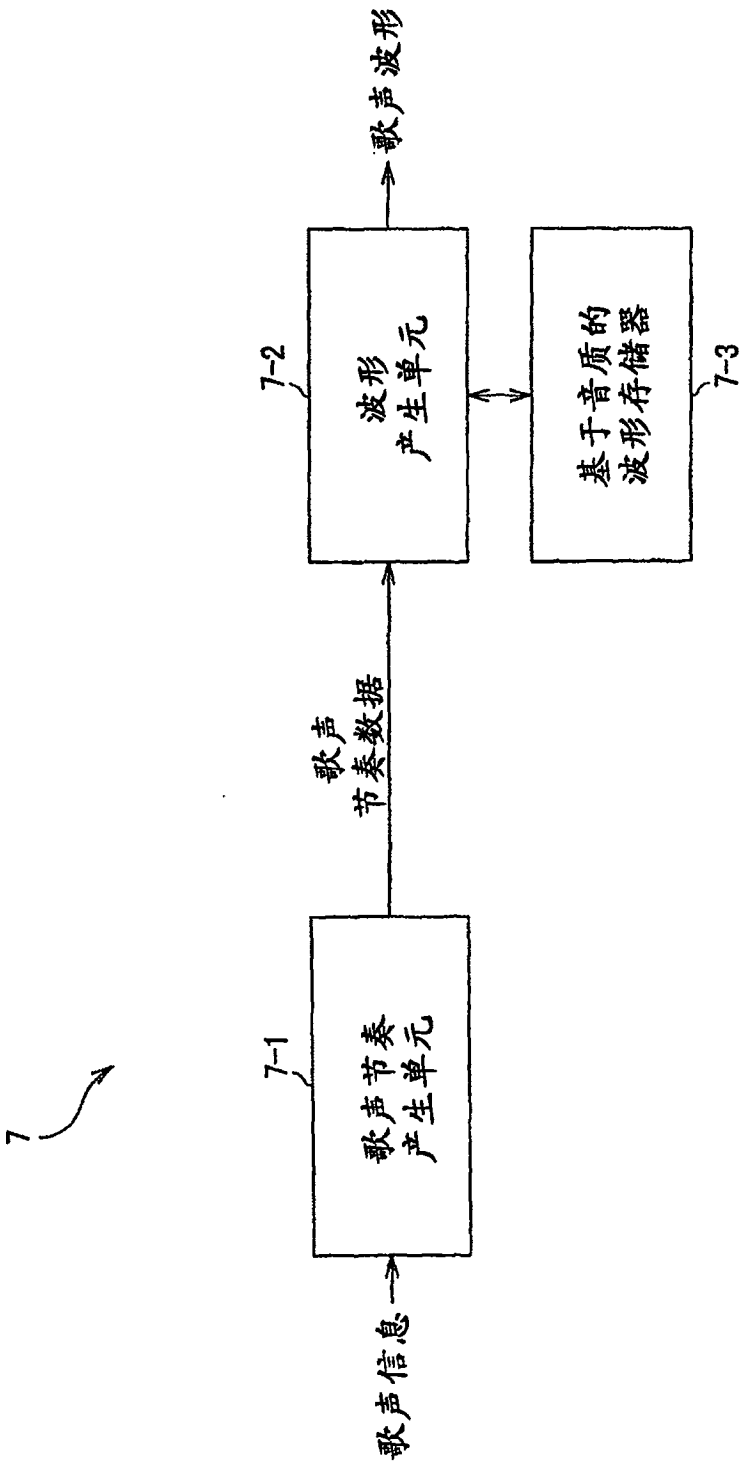


图 4

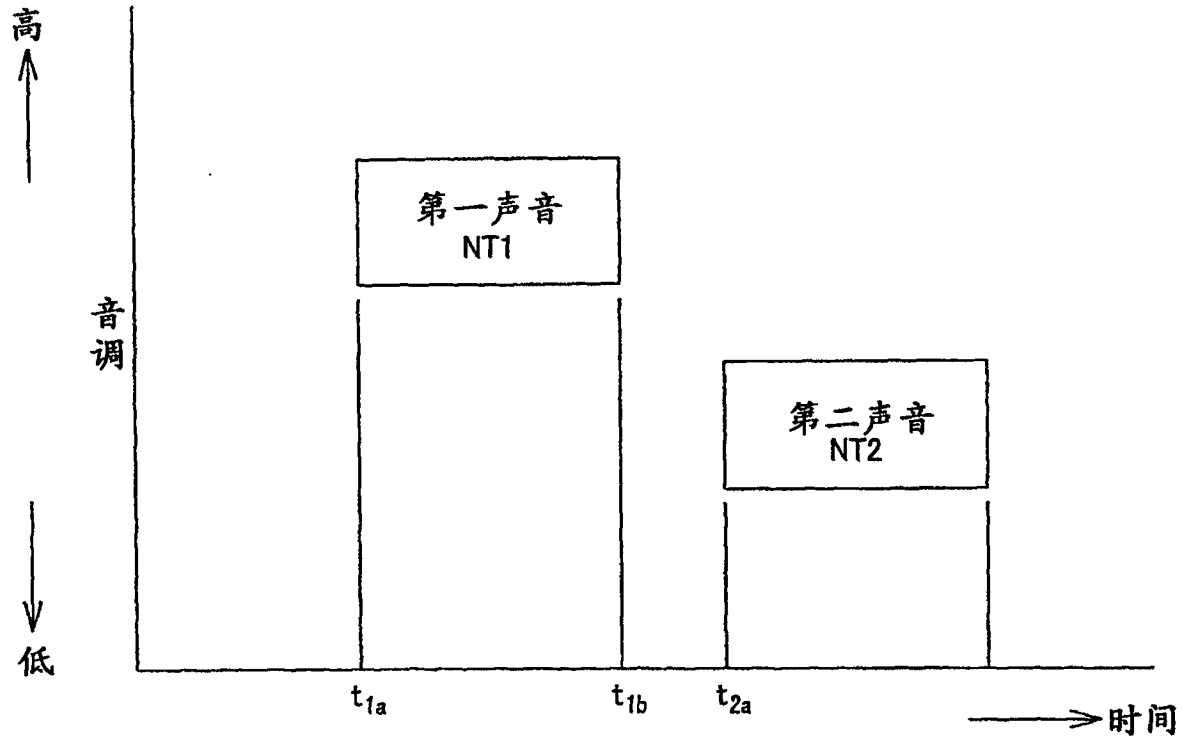


图 5

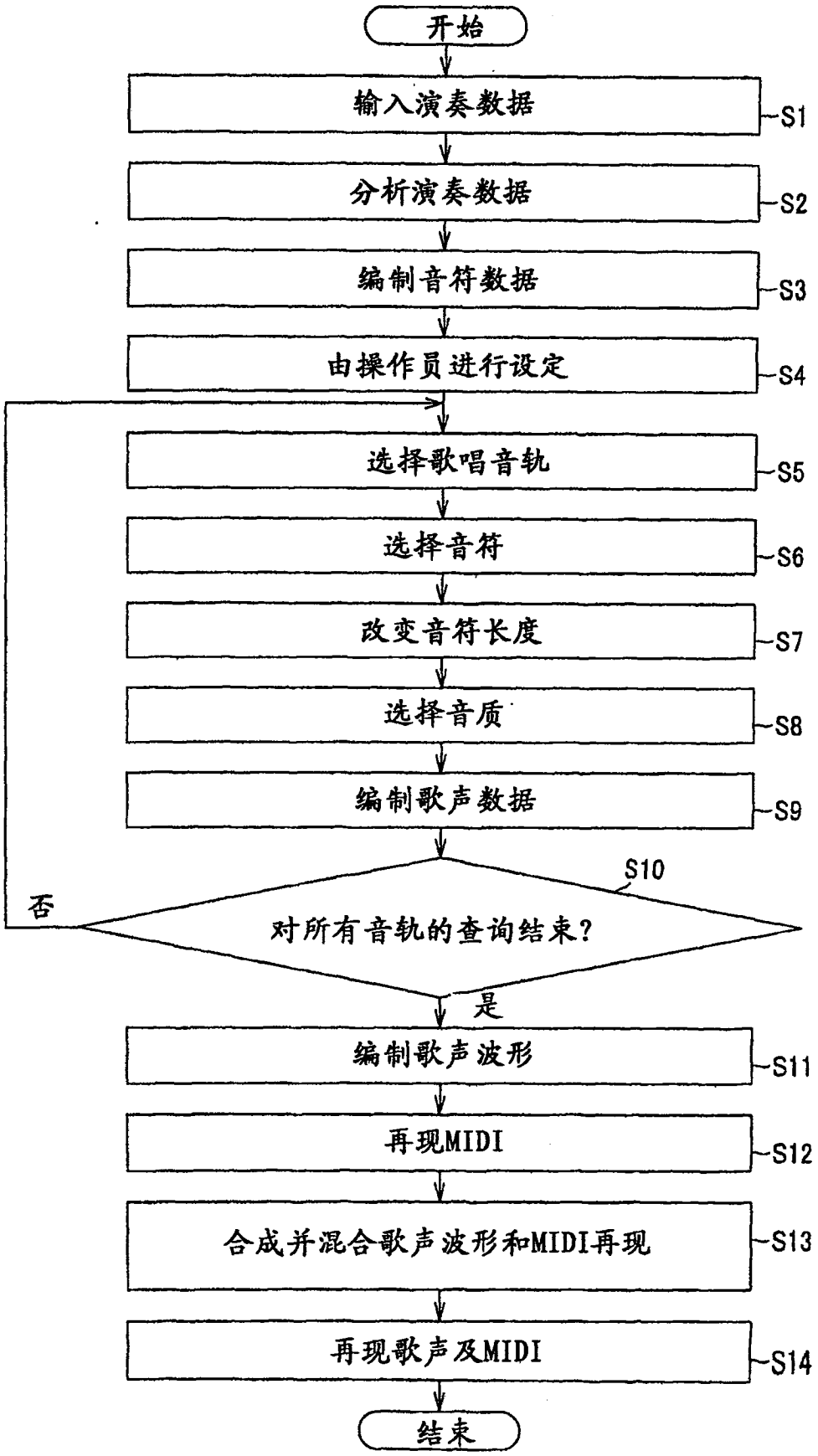

开始
输入演奏数据
S1
分析演奏数据
S2
编制音符数据
S3
由操作员进行设定
S4
选择歌唱音轨
S5
选择音符
S6
改变音符长度
S7
选择音质
S8
编制歌声数据
S9
S10
否
对所有音轨的查询结束?
是
编制歌声波形
S11
再现MIDI
S12
合成并混合歌声波形和MIDI再现
S13
再现歌声及MIDI
S14
结束

图 6

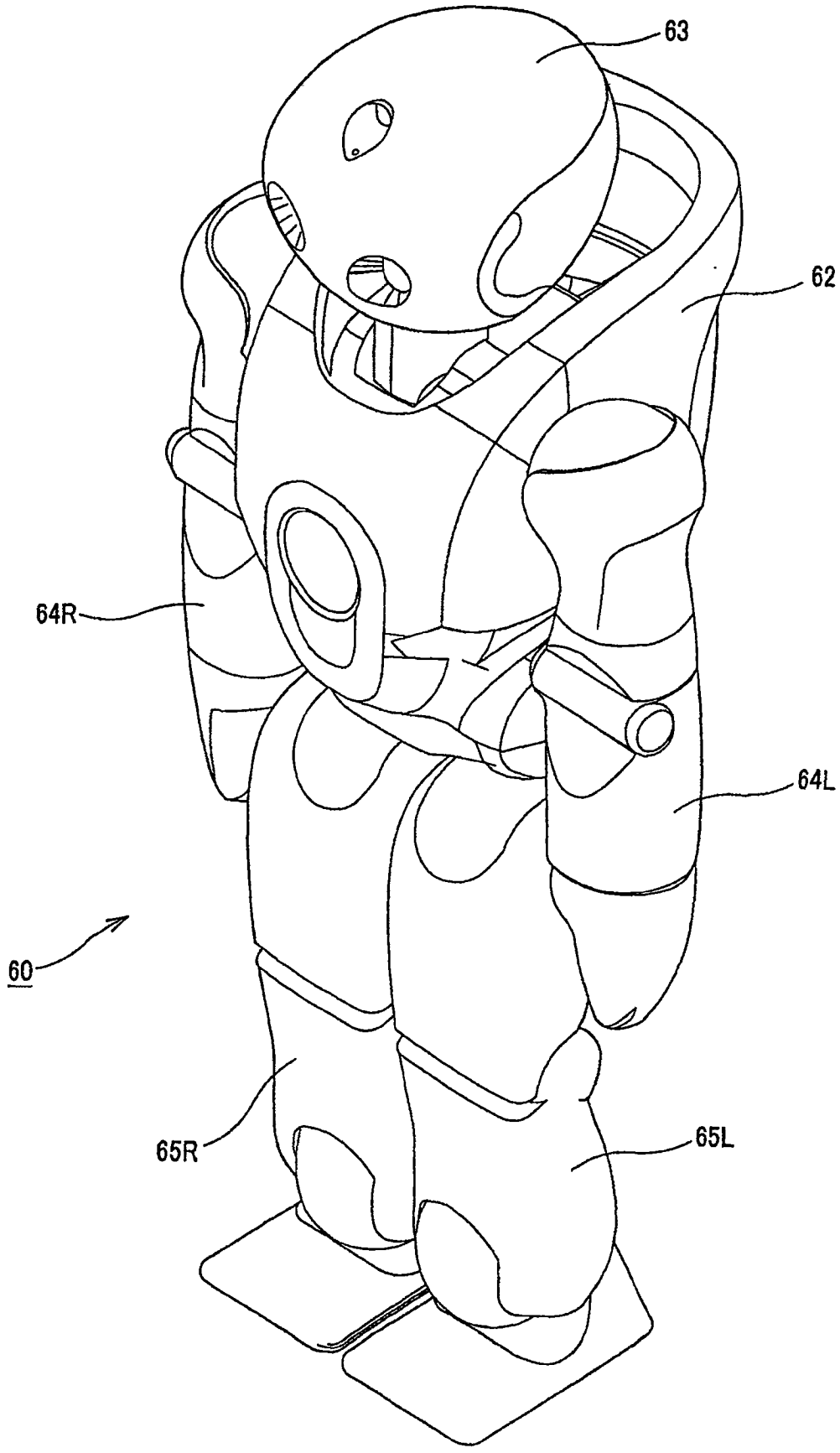


图 7

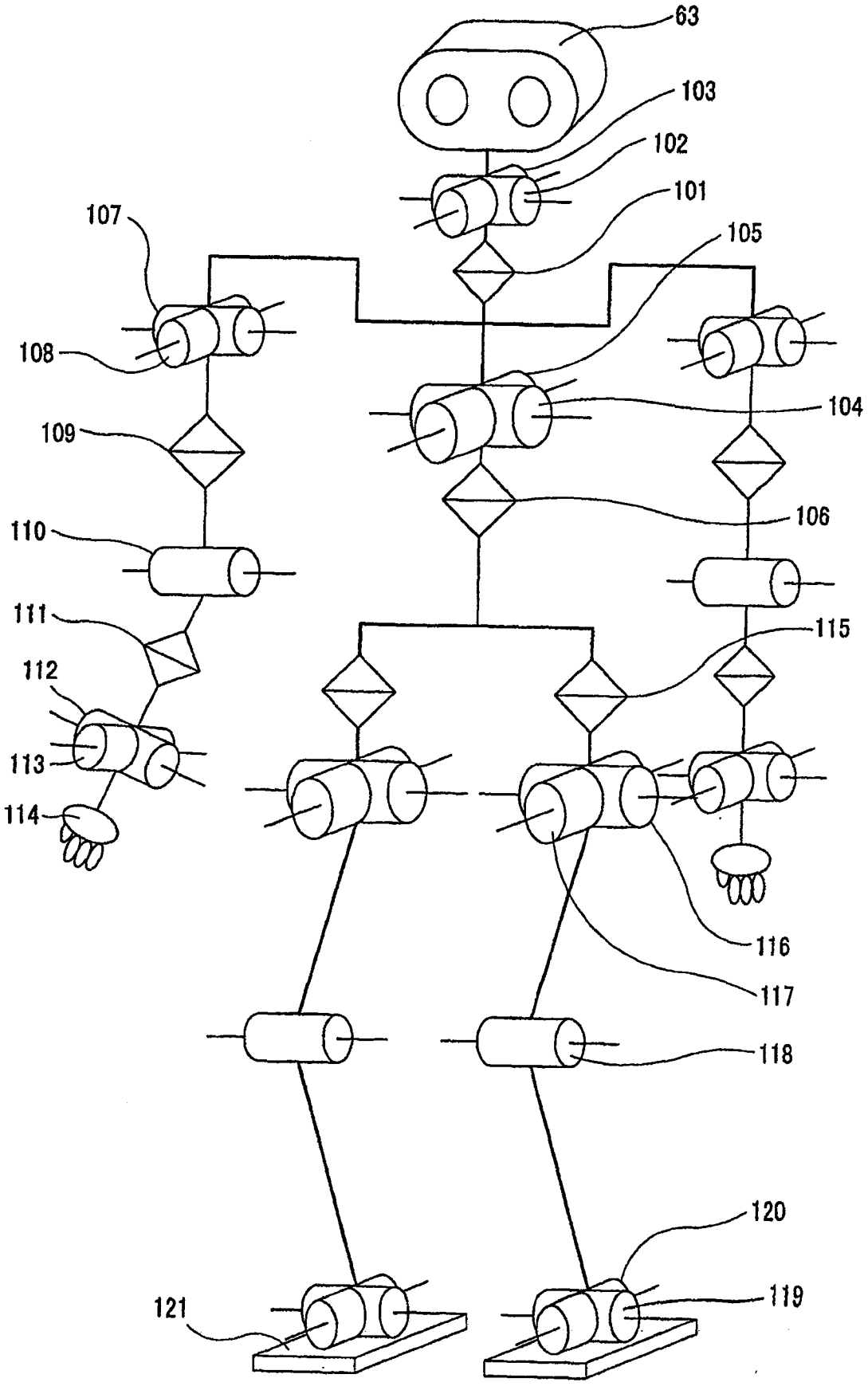

63
103
102
101
107
105
108
104
109
106
110
115
111
112
113
116
114
117
118
120
121
119

图 8

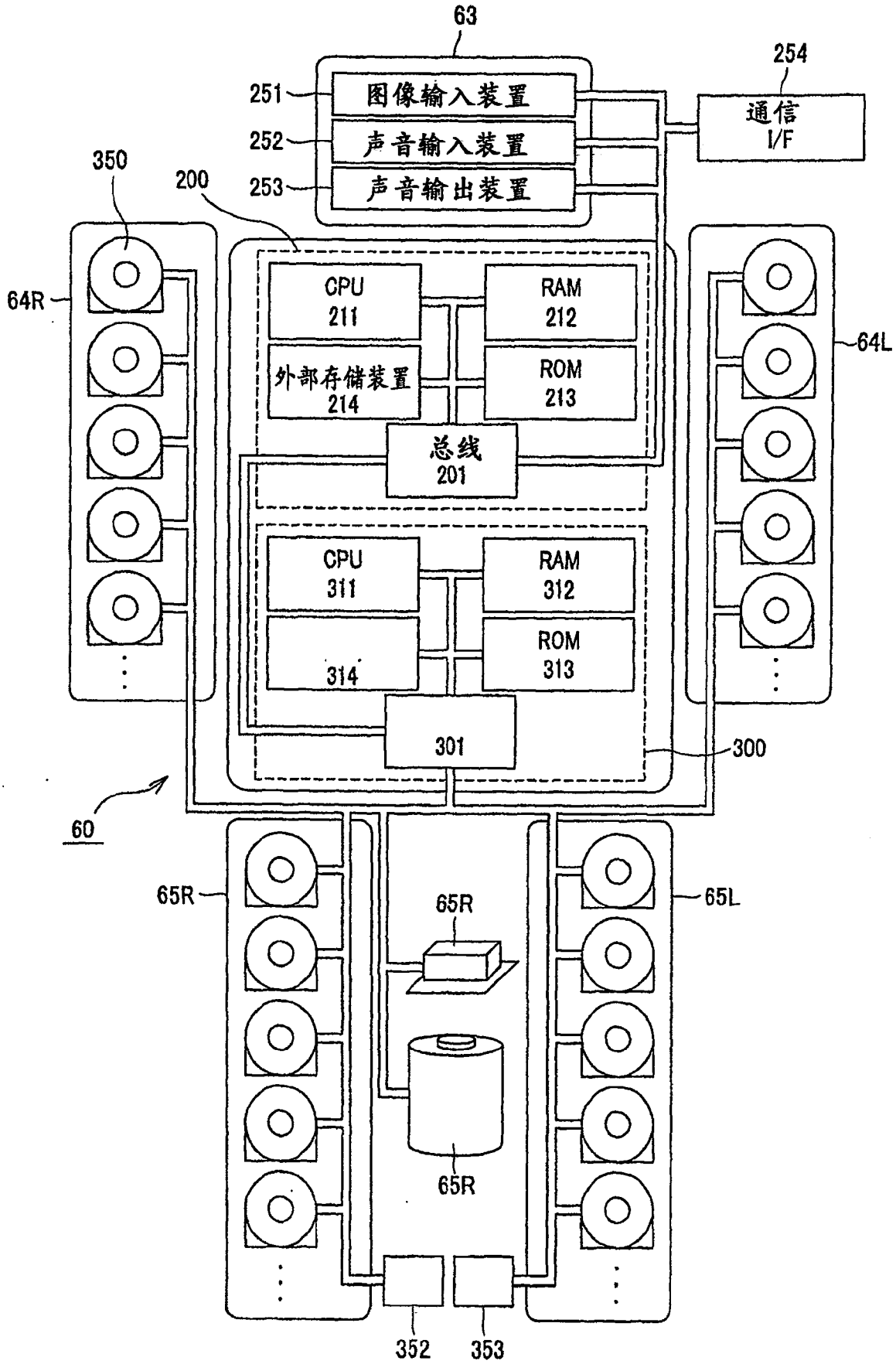

63
254
251
图像输入装置
252
声音输入装置
253
声音输出装置
通信
I/F
350
200
64R
64L
CPU
211
RAM
212
外部存储装置
214
ROM
213
总线
201
CPU
311
RAM
312
314
ROM
313
301
300
60
65R
65R
65L
65R
352
353

图 9