



(12)发明专利

(10)授权公告号 CN 105247609 B

(45)授权公告日 2019.04.12

(21)申请号 201480031099.X

(22)申请日 2014.06.02

(65)同一申请的已公布的文献号
申请公布号 CN 105247609 A

(43)申请公布日 2016.01.13

(30)优先权数据

2013-115111 2013.05.31 JP

2013-198217 2013.09.25 JP

2013-198218 2013.09.25 JP

2013-198219 2013.09.25 JP

2013-203839 2013.09.30 JP

2013-203840 2013.09.30 JP

2013-205260 2013.09.30 JP

2013-205261 2013.09.30 JP

2014-048636 2014.03.12 JP

(85)PCT国际申请进入国家阶段日
2015.11.30

(86)PCT国际申请的申请数据
PCT/JP2014/064631 2014.06.02

(87)PCT国际申请的公布数据
W02014/192959 JA 2014.12.04

(73)专利权人 雅马哈株式会社
地址 日本静冈县

(72)发明人 松原弘明 浦纯也 川原毅彦
久凑裕司 吉村克二

(74)专利代理机构 北京天昊联合知识产权代理
有限公司 11112

代理人 李铭 崔利梅

(51)Int.Cl.

G10L 13/00(2006.01)

G10L 13/033(2006.01)

G10L 13/10(2006.01)

G10L 15/22(2006.01)

(56)对比文件

JP 特开2003-271194 A,2003.09.25,

JP 特开2003-271194 A,2003.09.25,

US 2002198707 A1,2002.12.26,

CN 101981613 A,2011.02.23,

JP 特开2001-242881 A,2001.09.07,

CN 1725792 A,2006.01.25,

CN 1592330 A,2005.03.09,

CN 1781140 A,2006.05.31,

CN 1173086 A,1998.02.11,

CN 1672178 A,2005.09.21,

审查员 董小东

权利要求书3页 说明书45页 附图41页

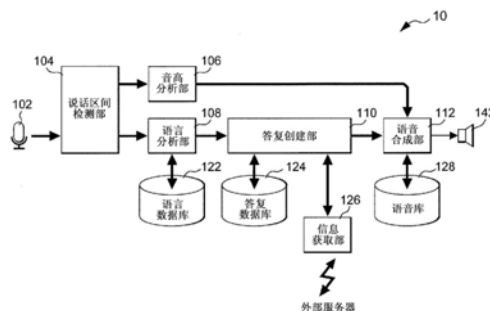
(54)发明名称

利用言语合成对话语进行响应的方法及装置

(57)摘要

本发明设置有:语音输入单元(102),其经由语音信号接收话语(询问);响应产生单元(110),其创建针对所述话语的响应(回答)的语音序列;音高分析单元(106),其对所述话语的第一区间(例如,句尾)的音高进行分析;以及语音产生单元(语音合成单元(112)等),其经由语音产生由所述语音序列表示的响应。所述语音产生单元以如下方式控制整个响应的音高:使所述响应的第二区间(例如,句尾)的音高相对于所述话语的第

一区间的音高实现规定的音高(例如,向下五度)。由此实现对能够给用户自然感觉的响应语音的合成。



1. 一种语音合成设备,包括:

语音输入部,其被构造为接收话语的语音信号;

音高分析部,其被构造为分析所述话语的第一区间的音高;

获取部,其被构造为获取所述话语的答复;和

语音产生部,其被构造为产生由所述获取部所获取的所述答复的语音,所述语音产生部以如下方式对所述答复的语音的整个语音波形数据执行音高转换来控制所述答复的语音的音高:使所述答复的第二区间具有与通过所述音高分析部分析的所述第一区间的音高相关联的音高,其中所述第一区间是作为询问的所述话语的词尾,而所述第二区间是所述答复的词首或词尾。

2. 根据权利要求1所述的语音合成设备,其中所述语音产生部以如下方式控制所述答复的语音的音高:使所述第二区间的音高相对于所述第一区间的音高的音程变为除完全一度以外的协和音程。

3. 根据权利要求1所述的语音合成设备,其中所述语音产生部以如下方式控制所述答复的语音的音高:使所述第二区间的音高相对于所述第一区间的音高的音程变为在从所述第一区间的音高向上一个八度和向下一个八度的范围内的除完全一度以外的任一音程。

4. 根据权利要求2或3所述的语音合成设备,其中所述语音产生部以如下方式控制所述答复的语音的音高:使所述第二区间的音高相对于所述第一区间的音高的音程变为从所述第一区间的音高向下五度的协和音程。

5. 根据权利要求2或3所述的语音合成设备,其中所述语音产生部将所述答复的语音的第二区间的音高临时性地设置在与所述第一区间的音高相关联的音高处,并且

所述语音产生部还被构造为执行以下操作中的至少一项:

在临时性设置的所述第二区间的音高低于预定的第一阈值的情况下,将临时性设置的所述第二区间的音高改变为向上移动一个八度的音高的操作;以及

在临时性设置的所述第二区间的音高高于预定的第二阈值的情况下,将临时性设置的所述第二区间的音高改变为向下移动一个八度的音高的操作。

6. 根据权利要求2或3所述的语音合成设备,其中所述语音产生部将所述答复的语音的第二区间的音高临时性地设置在与所述第一区间的音高相关联的音高处,并且

所述语音产生部还被构造为根据指定的属性将临时性设置的音高改变为向上或向下移动一个八度的音高。

7. 根据权利要求1所述的语音合成设备,其中第一模式和第二模式中的任一个可设置作为所述语音产生部的操作模式,并且

其中,在所述第一模式下,所述语音产生部以如下方式控制所述答复的语音的音高:使所述第二区间的音高相对于所述第一区间的音高的音程变为除完全一度以外的协和音程,以及

在所述第二模式下,所述语音产生部以如下方式控制所述答复的语音的音高:使所述第二区间的音高相对于所述第一区间的音高的音程变为不协和音程。

8. 根据权利要求1所述的语音合成设备,还包括:

非语言分析部,其分析与所述话语相关的、不同于音高的非语言信息;和

控制部,其根据所分析的非语言信息来控制所述语音产生部中对所述答复的语音产

生。

9. 根据权利要求1所述的语音合成设备,还包括:

语言分析部,其分析包括在所述话语和所述答复中的语言信息;以及

控制部,其根据所分析的语言信息来控制所述语音产生部中对所述答复的语音产生。

10. 根据权利要求1所述的语音合成设备,还包括:

非语言分析部,其分析所述话语中的音高变化;和

控制部,其控制在所述语音产生部中产生的所述答复的语音的音高根据所述话语中的音高变化而变化。

11. 根据权利要求1所述的语音合成设备,其中所述语音产生部被构造为根据给定的规则将所述第二区间的音高与所述第一区间的音高相关联并产生具有基于给定的中介属性的特点的语音,并且

所述语音合成设备还包括控制部,其基于所述中介属性和所述话语的讲话者的属性中的至少一个来确定所述规则。

12. 根据权利要求1所述的语音合成设备,其中所述语音产生部被构造为根据给定的规则将所述第二区间的音高与所述第一区间的音高相关联并产生具有基于给定的中介属性的特点的语音,并且

所述语音合成设备还包括控制部,其在所述答复的语音发声之后基于经由所述语音输入部接收的另一话语的语音来更新所述规则。

13. 一种计算机实现的方法,包括:

接收话语的语音信号;

分析所述话语的第一区间的音高;

获取所述话语的答复;

合成所获取的答复的语音;和

以如下方式对所述答复的语音的整个语音波形数据执行音高转换来控制所述答复的音高:使所述答复的语音的第二区间的音高具有与所分析的所述第一区间的音高相关联的音高,其中所述第一区间是作为询问的所述话语的词尾,而所述第二区间是所述答复的词首或词尾。

14. 一种编码/解码装置,包括:

A/D转换器,其将话语的输入语音信号转换为数字信号;

音高分析部,其基于所述数字信号来分析所述话语的第一区间的音高;

非正式反馈获取部,其在针对所述话语要返回非正式反馈时,获取与所述话语的含义相对应的非正式反馈数据;

音高控制部,其以如下方式对答复的语音的整个语音波形数据执行音高转换来控制所述非正式反馈数据的音高:使所述非正式反馈数据的第二区间具有与所分析的所述第一区间的音高相关联的音高,其中所述第一区间是作为询问的所述话语的词尾,而所述第二区间是所述答复的词首或词尾;和

D/A转换器,其被构造为将音高受控的非正式反馈数据转换为模拟信号。

15. 根据权利要求14所述的编码/解码装置,其中由所述A/D转换器转换的数字信号被供应至主机计算机,

所述音高控制部还被构造为接收从所述主机计算机返回的响应于所述话语的答复语音数据并且以如下方式控制所述答复语音数据的音高:使所接收到的答复语音数据的第三区间具有与所分析的所述第一区间的音高相关联的音高,并且

所述D/A转换器还被构造为将音高受控的答复语音数据转换为模拟信号。

16. 一种语音合成方法,包括:

用A/D转换器来将话语的输入语音信号转换为数字信号;

用处理器基于所述数字信号来分析所述话语的第一区间的音高;

当针对所述话语要返回非正式反馈时,用所述处理器来获取与所述话语的含义相对应的非正式反馈数据;

用所述处理器以如下方式对答复的语音的整个语音波形数据执行音高转换来控制所述非正式反馈数据的音高:使所述非正式反馈数据的第二区间具有与所分析的所述第一区间的音高相关联的音高,其中,所述第一区间是作为询问的所述话语的词尾,而所述第二区间是所述答复的词首或词尾;和

用D/A转换器将音高受控的非正式反馈数据转换为模拟信号。

利用言语合成对话语进行响应的方法及装置

技术领域

[0001] 本发明涉及言语或语音合成设备和系统,其响应于通过语音输入产生的话语、询问或说话,提供答复输出,还涉及与语音合成有关的编码/解码装置。

背景技术

[0002] 近年来,提出了以下语音合成技术。所提出的语音合成技术的示例包括:合成和输出与用户的说话音调和语音质量对应的语音,从而以更类似于人的方式产生语音的技术(例如,参见专利文献1);以及分析用户的语音以诊断该用户的心理和健康状态等的技术(例如,参见专利文献2)。

[0003] 近年来还提出了一种语音交互或对话系统,其通过在识别由用户输入的语音的同时以合成的语音输出由场景所指定的内容来实现与用户的语音交互(例如,参见专利文献3)。

[0004] 现有文献:

[0005] 专利文献1:日本专利申请特许公开No.2003-271194

[0006] 专利文献2:日本专利No.4495907

[0007] 专利文献3:日本专利No.4832097

[0008] 假设这样一种对话系统,其将上述语音合成技术和语音交互或对话系统结合,并响应于用户的语音给出的询问(用户的口头提问)而搜索数据并以合成的语音来输出回答或答复。但是,在这种情况下,将出现如下问题:通过语音合成输出的语音给用户不自然的感觉,更具体地,好似机器在说话的感觉。

发明内容

[0009] 鉴于上述问题,本发明的一个目的是以一种用于通过使用语音合成来响应询问或话语的技术实现对能够给用户自然感觉的响应或答复语音的合成。更具体地,本发明寻求提供一种技术,其能够容易且可控制地实现给用户良好印象的答复语音、给出不好印象的答复语音等。

[0010] 在研究用以合成对用户所给出的询问(或话语)进行答复的语音的人机系统时,本发明的发明人等首先考虑人与人之间实际进行的是何种对话,关注于非语言信息(即,不同于言语信息或语言信息的非言语信息),特别是对对话进行表征的音高(频率)。

[0011] 这里,考虑人与人之间的对话,其中一个人(下文为“人b”)对另一个人(下文为“人a”)给出的询问返回答复。通常,在这种情况下,当人a讲出询问时,不仅人a而且将要答复该询问的人b都对该询问中给定区间的音高印象深刻。在对该询问返回具有同意、赞同、肯定等意义的答复时,人b以如下方式讲出答复语音:答复中对答复进行表征的部分(诸如词尾或词首)的音高与(相对于)使人印象深刻的询问的音高呈现预定关系,更具体地,协和音程(consonant interval)关系。发明人等认为,由于给人a对他或她的询问留下印象的音高与对人b的答复进行表征的部分的音高处于上述关系,因此人a将对人b的答复有舒服

和舒适的良好印象。

[0012] 此外,自不存在语言的远古时代起的很长时间,人们彼此之间就已经进行了交流。据推测在这种环境下人类语音的音高和音量在人类交流中扮演了非常重要的角色。还据推测,虽然在已经发展了语言的现代基于语音音高的交流被遗忘了,但是自远古时代使用的“预定的音高关系”能够给予“某种舒服的”感觉,因为这种预定的音高关系已经铭记在人类的DNA中并流传至今。

[0013] 下面描述人与人之间的对话的具体示例。当人a用日语讲出询问“sou de sho?”(意思是“是不是?”)时,人a和人b通常在他们的记忆中记住询问的词尾处(在这里强烈的表明了提醒、确认等意义)的“sho”的音高。在这种状态下,如果人b想要肯定地答复该询问,则他或她以如下方式返回答复“a,hai”(日语罗马字,意思是“奥,是的”):对该答复进行表征的部分(例如,部分“i”(日语罗马字))的音高与记在人a和人b的记忆中的“sho”(日语罗马字)的音高呈现了上述关系。

[0014] 图2是示出在实际对话中的共振峰的示图,其中横轴表示时间而纵轴表示频率。更具体地,图2中的频谱指示了随着白色水平的增加的更大的强度。

[0015] 如图2所示,通过对人类语音的频率分析而获得的频谱显示为随时间移动的多个峰,即,显示为共振峰。更具体地,与“sou de sho?”(日语罗马字)对应的共振峰和与“a,hai”(日语罗马字)对应的共振峰均显示为三个峰区域(即,沿着时间轴移动的白色带状部分)

[0016] 现在参考这三个峰区域中最低频率的第一共振峰,在与“sou de sho?”(日语罗马字)中的“sho”对应的参考标记A处示出的频率(即,在参考标记A处示出的部分中的中央部分的频率)为约400 Hz,并且在与“a,hai”(日语罗马字)中的“i”对应的参考标记 B处示出的部分的频率为约260Hz。因此,可以看出在参考标记A 处示出的频率为在参考标记B处示出的频率的大约3/2。

[0017] 就音程而言,频率比为3/2的关系指的是从相同八度的音高“G”到音高“C”的音程,从音高“E”到从音高“E”向下(低)一个八度的音高“A”的音程等,也就是,如下所述的“完全五度”的关系。这种频率比(即,音高之间的预定的关系)只是一个优选示例,在本发明中还可使用各种其他频率比的示例。

[0018] 图3是示出音名(音节名)与人类语音的频率之间的关系的示图,其还示出了利用第四八度中的“C”作为基准音高时的频率比。“G”与基准音高“C”的频率比是如上所述的3/2。图3还示出了利用第三八度中的“A”作为基准音高时的频率比。

[0019] 即,可以认为在人与人之间的对话中,询问的音高和答复的音高处于前述预定的关系,而不是彼此不相关。此外,本发明的发明人等通过分析许多对话以及统计地集合许多人进行的评估,确认上述观察结果通常是正确的。在上述观察结果和证实的启发下,在对以合成的语音输出对用户讲出的询问的答复的对话系统进行研究时,本发明的发明人以如下方式设置语音合成以便实现本发明的上述目的。

[0020] 即,为了实现上述目的,本发明的一个方面提供了一种语音合成设备,其包括:语音输入部,其被构造为接收话语的语音信号;音高分析部,其被构造为分析话语的第一区间的音高;获取部,其被构造为获取话语的答复;以及语音产生部,其被构造为产生由获取部所获取的答复的语音,所述语音产生部以如下方式控制所述答复的语音的音高:所述答复

的第二区间具有与通过音高分析部分分析的第一区间的音高相关联的音高。

[0021] 根据本发明的这样的实施例,可以防止响应于询问(话语)的输入语音信号而合成的答复的语音伴有不自然的感觉。注意,对询问(话语)的答复不限于特定或具体的答复,并且有时可以是非正式反馈(back-channel feedback)(插入语)的形式,诸如“ee”(日语罗马字,意思是“呀。”)、“naruhodo”(“我明白了。”)或“sou desune”(“我同意。”)。此外,答复不限于人类语音答复,并且有时可以是动物的语音的形式,诸如“wan”(“汪”)或“Nyâ”(“喵”)。即,本文中使用的术语“答复”和“语音”指的是不仅包含人发出的语音而且包含动物的语音的概念。

[0022] 在本发明的实施例中,第一区间优选是话语(即,询问)的词尾,而第二区间优选是答复的词首或词尾。因为在很多情况下,表征询问的印象的部分是答复的词尾,而表征答复的印象的部分是答复的词首或词尾。

[0023] 在本发明的实施例中,语音产生部可以构造为以如下方式控制答复的语音的音高:使第二区间的音高相对于第一区间的音高的音程变为除完全一度以外的协和音程。这里,“协和”意即已同时产生的多个音调彼此混合以彼此良好地和谐的关系,并将这种音程关系称作协和音程。两个音符之间的频率比越简单,和谐度就变得越高。最简单的频率比1/1(完全一度)和频率比2/1(完全八度)被称作绝对协和音程,而频率比3/2(完全五度)和频率比4/3(完全四度)加上上述的频率比1/1(完全一度)和频率比2/1(完全八度)被称作完全协和音程。此外,频率比5/4(大三度)、6/5(小三度)、5/3(大六度)和8/5(小六度)被称作不完全协和音程。此外,除上述之外的所有其他频率比(诸如大二度、小二度、大七度、小七度和各种增音程和减音程)被称作不协和音程。

[0024] 因为预料到在答复的词首或词尾具有与询问的词尾相同音高的情况下会涉及到作为对话的不自然感觉,因此从询问的音高和答复的音高之间的关系中排除完全同度或完全一度。

[0025] 在本发明的实施例中,询问的音高和答复的音高之间的音程不限于除完全一度以外的协和音程,并且可以是如下预定范围内的任一音程。即,语音产生部可构造为以如下方式控制答复的语音的音高:使第二区间的音高相对于第一区间的音高的音程变为预定范围内除完全一度以外的任一音程,该预定范围为从第一区间的音高向上和向下一个八度。如果答复的音高从询问的音高向上或向下移动一个八度,则不仅无法建立协和音程关系,而且对话将变得不自然。同样在本实施例中,将完全一度从询问的音高向上和向下一个八度的范围内的音高关系中排除是因为如上所述在答复的音高和询问的音高相同的情况下对话可能变得不自然。

[0026] 在本发明的实施例中,语音产生部可以构造为以如下方式控制答复的语音的音高:使第二区间的音高相对于第一区间的音高的音程变为从第一区间的音高向下五度的协和音程。该设置使得讲出询问的用户对返回给询问的答复具有良好印象。

[0027] 此外,在本发明的实施例中,语音产生部可构造为将答复的语音的第二区间的音高临时性地设置在与第一区间的音高相关联的音高处,并且语音产生部可以进一步构造为执行以下操作的至少一项:在临时性设置的第二区间的音高低于预定的第一阈值的情况下,将临时性设置的音高改变为向上移动一个八度的音高的操作;在临时性设置的第二区间的音高高于预定的第二阈值的情况下,将临时性设置的音高改变为向下移动一个八度的

音高的操作。因为在临时性设置的第二区间的音高低于预定的第一阈值(或高于预定的第二阈值)的情况下,语音产生部将临时性设置的第二区间的音高移动至向上(或向下)一个八度的音高,所以本实施例可以防止以不自然地过低的音高(或不自然地过高的音高)合成答复的语音。

[0028] 在本发明的实施例中,语音产生部可以构造为将答复的语音的第二区间的音高临时性地设置在与第一区间的音高相关联的音高处,并且语音产生部可以进一步构造为根据指定的属性将临时性设置的音高改变为向上或向下移动一个八度的音高。这里,“属性”例如是要合成的语音的属性,并且属性的示例包括女性、儿童和成年男性等的属性。同样在这种情况下,语音产生部将临时性设置的第二区间的音高移动至向上(或向下)一个八度的音高,并且由此,本实施例可以鉴于指定的属性来防止以不自然地过低的音高(或不自然地过高的音高)合成答复的语音。

[0029] 此外,在本发明的实施例中,第一模式和第二模式中的任一个可以设置作为语音产生部的操作模式。在第一模式中,语音产生部可以构造为以如下方式控制答复的语音的音高:使第二区间的音高相对于第一区间的音高的音程变为距第一区间的音高的除完全一度之外的协和音程。在第二模式下,语音产生部可以构造为以如下方式控制答复的语音的音高:使第二区间的音高相对于第一区间的音高的音程变为距第一区间的音高的不协和音程。在第二模式下,与询问(话语)处于不协和音程关系的语音被合成,并且由此,本实施例可向讲出询问的用户给出不舒服的感觉。反而言之,通过将语音产生部的操作模式设置为第二模式,本实施例可以吸引用户的注意或者有意地向用户给出威胁感。另一方面,通过将语音产生部的操作模式设置为第一模式,本实施例允许询问(话语)和答复(响应)之间的协和音程关系,并且由此能够给用户更好的感觉。由此,通过允许根据情况来恰当使用第一模式和第二模式中的任一个,本实施例可以显著增强语音合成设备的使用性。

[0030] 当语音产生部的操作模式是第一模式时,可以按如下方式执行控制:在存在预定时间范围内的“停顿”(或时间间隔)的情况下输出处于除了完全一度之外的协和音程关系的答复,在这种情况下可以防止给讲出询问(话语)的用户不舒服的感觉。这里,0.5秒至2.0秒的时间范围内的停顿被认为是恰当的。

[0031] 通常,在人与人之间的对话中,询问的音高是对询问的印象进行表征的因素,但是询问的印象还大大地受到不同于音高的非语言信息的影响。当人小声地如低语般讲出询问时,给另一方这样的印象:暗含要求小声讲出对该询问的答复。此外,即使当一个人在他或她的面部带有悲伤表情(悲伤的面部表情)的情况下说出“我不伤心”时,这个人也被认为是非言语地悲伤,尽管这个人可能听起来言语地悲伤,因此,答复的印象必须根据话语的非言语(非口头)含义来做出。因此,在研究以合成的语音输出对用户讲出的询问的答复的对话系统时,不仅是询问的音高,而且还有不同于音高的非言语信息可成为对答复的语音进行合成时的重要因素。

[0032] 因此,在实施例中,本发明的语音合成设备可以进一步包括:非语言分析部,其分析与话语相关的、不同于音高的非语言信息;以及控制部,其根据所分析的非语言信息来控制语音产生部中对答复的语音产生。因为控制部根据与话语相关的、不同于音高的非语言信息(例如,包括在话语中的、不同于音高的非语言信息,诸如音量和速度、或者讲出话语的用户的面部表情)来控制答复的语音(答复语音),因此可以合成与(与话语相关的、不同

于音高的)非语言信息相和谐的答复语音。因此,本发明可以合成带有不自然感觉的答复语音。

[0033] 注意,要根据本发明合成或产生的答复可以是非正式反馈(插入语)(诸如“ee”、“naruhodo”或“soudesune”),而非限于对询问(话语)的具体答复。此外,除了类似于音量、语音质量和速度(讲话速度)的与询问的语音相关的信息外,询问中的不同于音高的非语言(非言语)信息的示例还包括讲出询问的用户的面部表情、姿势、衣物等。

[0034] 附带提及,既然如上所述在人与人之间的对话中询问的音高是表征该询问的印象的因素,则按照经验认为,答复(响应)的方式根据询问和答复(响应)的言语或语言内容而不同。例如,在要返回答复“是(hai)”的情况下,对询问的答复相对快速地返回,但是,在要返回答复“不(iie)”的情况下,考虑到将要给对话中的另一方的印象,对询问的答复略微停顿(时间间隔)地返回。因此,在研究以合成的语音输出对用户讲出的询问(话语)的答复的对话系统时,询问(话语)和答复的言语或语言意义也可以成为在合成答复的语音时的重要因素。

[0035] 鉴于以上情况,本发明的实施例还包括:语言分析部,其分析包括在话语和答复中的语言信息;以及控制部,其根据所分析的语言信息来控制语音产生部中对答复的语音产生。因为答复语音的合成是根据包括在话语(询问)和答复(响应)中的语言信息来控制的,因此尽管答复语音是机器合成的语音,但也好似正在与一个人进行对话的印象。注意,上述根据包括在答复(响应)中的语言信息对该答复(响应)自身的语音的合成进行控制的一个示例形式可以包括根据该答复(响应)具有肯定含义还是否定含义来对该答复(响应)的语音进行控制。

[0036] 顺带提及,已经验性地认识到,答复(响应)的方式取决于例如话语中的音高随时间变化的方式而不是仅该音高本身而不同。例如,即使讲出的话语以名词结尾,比如,“asu wa hare”(日语罗马字,意思是“明天是晴天”),如果音高朝向词尾上升的话,其也会变成意思是“asu wa hare desuka?”(“明天是晴天吗?”)的询问(疑问句)。如果在话语“asu wa hare”中音高大致恒定,则该话语应当为某种纯粹的独白或自言自语。因此,响应于该话语的比如“so desune”之类的答复(非正式反馈)也将是音高大致恒定。因此,在研究以合成的语音返回对用户讲出的话语的答复的对话系统时,不仅是话语的音高,而且还有指示话语的音高变化方式的非言语或非语言信息可成为对答复的语音进行合成时的重要因素。

[0037] 鉴于以上情况,本发明的一个实施例还可以包括:非语言分析部,其分析话语中的音高变化;以及控制部,其控制在语音产生部中产生的答复的语音的音高根据话语中的音高变化而变化。因为根据话语中随时间的音高变化来控制答复中的音高变化,因此即使答复的语音(答复语音)是机器合成的语音,也可以给用户好似正与人进行对话的印象。注意,上述根据话语中音高变化的方式对答复的语音的音高进行控制的一个示例形式可以包括当话语中几乎不存在音高随时间变化时(即,当话语中的音高是平坦的时)使作为非正式反馈的答复为平坦的。此外,如果话语是疑问句(其中音高朝向词尾上升),则根据话语中音高变化的方式对答复的语音的音高进行控制的一个示例形式可以包括使答复的音高朝向词尾降低。

[0038] 随便提及,在研究以合成的语音返回对用户讲出的话语的答复的对话系统时,可预料的是,各类人,不管性别和年龄,都成为该对话系统的用户。此外,存在要用于语音合成

的数据(比如语音片段)的典型模型。反而言之,只要提前准备多个模型用于语音合成,就可以合成各种语音质量的答复的语音。因此,当将通过语音合成(即,以合成的语音)来输出答复时,可以输出各种属性(中介属性(agent attribute))的答复。因此,在对话系统中,必须考虑存在用户的属性(说话者或讲话者的属性)和中介属性的各种组合。更具体地,在话语的说话者或讲话者是女性而答复者是男性的情况下,并且如果该男性试图以如下方式对所述话语进行答复:使针对所述话语的答复的词尾等的音高与所述女性的话语的词尾的音高呈现出预定的关系,则所述答复的词尾等的音高对于该男性而言将过高,使得该答复将不期望地变得不自然。相反,在话语的说话者或讲话者是男性而答复者是女性的情况下,并且如果该女性试图以如下方式对所述话语进行答复:使所述话语的答复的词尾等的音高与所述男性的话语的词尾的音高呈现出预定的关系,则所述答复的词尾等的音高对于该女性而言将过低。

[0039] 鉴于上述情况,提出了本发明的一个实施例,其中所述语音产生部可以被构造为根据给定的规则将所述第二区间的音高与所述第一区间的音高相关联并产生具有基于给定的中介属性的特点的语音,并且该实施例还可以包括控制部,其基于所述中介属性和所述话语的讲话者的属性中的至少一个确定所述规则。

[0040] 根据本实施例,用于将所述第二区间的音高与所述第一区间的音高相关联的所述规则基于所述话语的讲话者的属性或所述中介属性来确定。因此,音高根据所确定的规则而受控制的所述答复的语音的音高特点将根据所述话语的讲话者的属性和所述中介属性中的至少一个而可变地受控制。因此,即使针对话语的答复的语音是机器合成的语音,也可以给用户自然的对话感觉,并且利用本发明的语音合成设备进行对话可以给用户某种程度的愉悦感觉。

[0041] 所述中介属性是要用于合成语音的模型的属性,诸如性别或年龄。作为一个示例,在所述语音合成设备中预设一个或多个中介属性,使得可以酌情选择或设置任一中介属性来用于语音的合成。此外,所述话语的讲话者的属性是该讲话者的性别,比如男性、女性或中性。此外,除了性别以外,所述话语的讲话者的属性的示例还可以包括年龄和一代人(比如儿童、成年人和老年人)。可以在话语被输入到语音合成设备时提前将这种讲话者的属性指示给语音合成设备,或者可以由语音合成设备对所接收到的话语的语音信号进行分析而自动地识别这种讲话者的属性。

[0042] 即使尝试基于统计来找出可使人们感到舒服等的音高关系,但是这种可使人们感到舒服等的音高关系也因人而异,因此,仅通过统计分析会对获取更好的规则造成限制。此外,在以合成的语音对特定用户讲出的话语返回答复的对话系统中,重要的是增加该特定用户的话语的数量和频率,或者简言之,使该特定用户与机器之间的对话有生气。

[0043] 鉴于上述情况,提出了本发明的一个实施例,其中所述语音产生部可被构造为根据给定的规则将所述第二区间的音高与所分析的所述第一区间的音高相关联并产生具有基于给定的中介属性的特点的语音,并且该实施例还可以包括控制部,其在使所述答复的语音发声之后基于经由所述语音输入部接收到另一话语的语音来更新所述规则。用于将所述答复的第二区间的音高与所分析的所述第一区间的音高相关联的上述规则在使所述答复的语音产生或发声之后基于经由所述语音输入部接收到另一话语的语音而更新。即,要应用于下一个答复的语音的发声或产生的规则在考虑了用户响应于机器合成的答复而进

一步讲出的话语的情况下酌情更新。这种规则更新特点可以使对话导向更逼真的方向。

[0044] 可以构造上述实施例使得上述规则根据多个预设场景中的任一个而设置。这里,所述场景的示例包括讲话者的性别和年龄与要合成的语音的性别和年龄的组合、话语的速度(快速说话和缓慢说话)与要合成其语音的答复的速度的组合、对话的目的(诸如语音指导)等。

[0045] 在人与人之间的典型对话中,其中一个人(人a)讲出话语,而另一个人(人b)响应于人a的话语而讲出话语。然而,人b响应于人a的话语的话语并不总是具体语句形式的答复,有时候也可能只不过是正式反馈的形式,比如“a,hai”、“sou desune”、“sorede”等。此外,凭经验已知的是,在人与人之间的对话中,这种非正式反馈以良好响应的方式快速地返回至人a的话语。因此,在研究以合成语音对用户讲出的话语返回答复的对话系统时,重要的是使得非正式反馈以良好响应的方式快速地作为简单答复输出。

[0046] 鉴于上述情况,根据本发明另一方面的一种编码/解码装置包括:A/D转换器,其将话语的输入语音信号转换为数字信号;音高分析部,其基于所述数字信号来分析所述话语的第一区间的音高;非正式反馈获取部,其在针对所述话语要返回非正式反馈时,获取与所述话语的含义相对应的非正式反馈数据;音高控制部,其以如下方式控制所述非正式反馈数据的音高:使所述非正式反馈数据的第二区间具有与所分析的所述第一区间的音高相关联的音高;以及D/A转换器,其被构造为将音高受控的非正式反馈数据转换为模拟信号。这种配置可以提供一种紧凑的编码/解码装置,其能够实现人机对话功能并且其以如下方式进行构造:作为模拟信号输入的话语在被转换为数字信号后进行处理,与话语的含义相对应的非正式反馈的语音数据被数字地创建,并且最终将模拟的非正式反馈语音信号输出。以这种方式,可以以良好响应的方式快速地创建的非正式反馈,并且可以防止不自然的感受伴随非正式反馈的合成语音。

[0047] 根据本发明的又一方面,提供了一种包括编码/解码装置和主机计算机的语音合成系统。这里,所述编码/解码装置包括:A/D转换器,其将话语的输入语音信号转换为数字信号;音高分析部,其基于所述数字信号来分析所述话语的第一区间的音高;非正式反馈获取部,其在针对所述话语要返回非正式反馈时,获取与所述话语的含义相对应的非正式反馈数据;音高控制部,其以如下方式控制所述非正式反馈数据的音高:使所述非正式反馈数据的第二区间具有与所分析的所述第一区间的音高相关联的音高;以及D/A转换器,其被构造为将音高受控的非正式反馈数据转换为模拟信号。所述主机计算机以如下方式构造:当针对所述话语要返回不同于所述非正式反馈的答复语音时,所述主机计算机根据由所述A/D转换器转换的数字信号获取响应于所述话语的答复语音数据并且将所获取的答复语音数据返回至所述编码/解码装置,所述音高控制部还被构造为以如下方式控制所述答复语音数据的音高:使所接收的答复语音数据的第三区间具有与所分析的所述第一区间的音高相关联的音高,并且所述D/A转换器还被构造为将音高受控的答复语音数据转换为模拟信号。

[0048] 因此,可以通过编码/解码装置进行的快速处理来高效地产生响应于输入话语的非正式反馈(简单答复)的语音,同时可以通过主机计算机进行的大容量集中式处理来高效地产生响应于输入话语的具有相对复杂的语言含义的答复的语音,从而提供一种能够实现灵活的人机对话功能的语音合成系统。

[0049] 本发明不仅可以具体化为上述语音合成设备或系统或者编码/解码装置,而且可

以具体化为计算机或处理器实现的方法。此外,本发明可以具体化为存储了包括指令集的软件程序的非暂时性计算机可读存储介质,所述指令集用于使计算机或处理器作为语音合成设备工作。

[0050] 在本发明中,询问的音高(频率)是分析的对象,而答复的音高是控制的对象,如上下文中所述。然而,根据以上讨论的共振峰的示例显而易见的是,人类语音具有确定的频带,因此,在分析和控制时,人类语音将不可避免地在确定频率范围内出现误差。此外,在分析和控制中不可避免地出现误差。因此,根据本发明,针对分析和控制,允许在确定范围内人类语音伴随有误差,而不是音高(频率)值始终恒定。

附图说明

- [0051] 图1是示出本发明的语音合成设备的第一实施例的构造的框图;
- [0052] 图2是示出对话中的语音的共振峰的示例的图;
- [0053] 图3是示出音名和频率等之间的关系的图;
- [0054] 图4是示出语音合成设备的第一实施例所执行的处理的流程图;
- [0055] 图5是示出识别出词尾的示例具体方式的图;
- [0056] 图6是示出对语音序列执行的音高移动操作的示例的图;
- [0057] 图7是示出语音合成对用户的话语给出的心理印象的图;
- [0058] 图8是示出本发明的语音合成设备的第二实施例的构造的框图;
- [0059] 图9是示出对语音波形数据执行音高转换的示例方式的图;
- [0060] 图10是示出应用示例1中的处理的主要部分的流程图;
- [0061] 图11是示出应用示例2中的处理的主要部分的流程图;
- [0062] 图12是示出应用示例3中的处理的主要部分的流程图;
- [0063] 图13是示出应用示例4中的处理的主要部分的流程图;
- [0064] 图14是示出语音合成设备的第三实施例的构造的框图;
- [0065] 图15是示出在语音合成设备的第三实施例中执行的处理的流程图;
- [0066] 图16是示出本发明的语音合成设备的第四实施例的构造的框图;
- [0067] 图17是对本发明的第五实施例中的语音合成进行说明的图;
- [0068] 图18是示出本发明的语音合成设备的第六实施例中执行的处理的主要部分的流程图;
- [0069] 图19是示出本发明的语音合成设备的第七实施例的构造的框图;
- [0070] 图20是示出本发明的语音合成设备的第八实施例的构造的框图;
- [0071] 图21是示出由语音合成设备的第八实施例执行的操作序列的流程图;
- [0072] 图22是示出对语音序列执行的音高移动操作的示例的图;
- [0073] 图23是示出对语音序列执行的音高移动操作的示例的图;
- [0074] 图24是示出对语音序列执行的音高移动操作的示例的图;
- [0075] 图25是对答复的输出时机进行说明的图;
- [0076] 图26是对答复的输出时机进行说明的图;
- [0077] 图27是示出本发明的语音合成设备的第九实施例的构造的框图;
- [0078] 图28是语音合成设备的第九实施例中执行的语音合成处理的流程图;

- [0079] 图29是示出识别出词尾和音高变化的特定具体示例方式的图；
- [0080] 图30是示出对语音序列执行的音高移动操作的示例的图；
- [0081] 图31是示出对语音序列执行的音高移动操作的示例的图；
- [0082] 图32是示出本发明的语音合成设备的修改例的构造的框图；
- [0083] 图33是示出本发明的语音合成设备的第十实施例的构造的框图；
- [0084] 图34是示出语音合成设备的第十实施例中执行的语音合成处理的流程图；
- [0085] 图35是示出图34的语音合成处理中的规则确定处理的细节的流程图；
- [0086] 图36是示出对语音序列执行的音高移动操作的示例的图；
- [0087] 图37是示出对语音序列执行的音高移动操作的示例的图；
- [0088] 图38是示出对语音序列执行的音高移动操作的示例的图；
- [0089] 图39是示出对语音序列执行的音高移动操作的示例的图；
- [0090] 图40是示出对语音序列执行的音高移动操作的示例的图；
- [0091] 图41是示出本发明的语音合成设备的第十一实施例的构造的框图；
- [0092] 图42是示出语音合成设备的第十一实施例中的指标表的示例的图；
- [0093] 图43是示出语音合成设备的第十一实施例中的操作期间的示例切换的图；
- [0094] 图44是示出语音合成设备的第十一实施例中执行的语音合成处理的流程图；
- [0095] 图45是示出语音合成设备的第十一实施例中执行的表更新处理的流程图；
- [0096] 图46是示出语音合成设备的第十二实施例中的指标表的示例的图；
- [0097] 图47是示出语音合成设备的第十三实施例中的指标表的示例的图；
- [0098] 图48是示出根据本发明的第十四实施例构造的语音合成设备的硬件设置的图；
- [0099] 图49是示出根据第十四实施例的语音合成设备的功能配置的功能框图；
- [0100] 图50是图49的语音合成设备中执行的语音处理的流程图；
- [0101] 图51是示出图50的语音处理中的非正式反馈处理的细节的流程图；以及
- [0102] 图52是示出图50的语音处理中的答复处理的细节的流程图。

具体实施方式

[0103] 现在,将在下文中参照附图对本发明的优选实施例进行详细描述。

[0104] <第一实施例>

[0105] 首先,将描述本发明的语音合成设备的第一实施例。图1是示出本发明的语音合成设备10的第一实施例的构造的框图。在图1中,语音合成设备10是终端设备(比如移动或便携式设备),其包括CPU(中央处理单元)、语音输入部102和扬声器142。在语音合成设备10中,通过CPU对预先安装的应用程序进行执行而按如下建立多个功能块。

[0106] 更具体地,在语音合成设备10中,建立了说话区间检测部104、音高分析部106、语言分析部108、答复创建部110、语音合成部112、语言数据库122、答复数据库124、信息获取部126和语音库128。即,通过与功能块相对应的程序模块和CPU之间的协作来实现所示语音合成设备10中的各功能块。

[0107] 虽然未特别示出,但是所述语音合成设备10还包括显示部、操作输入部等,使得用户能够检查该设备的状态并将各种操作输入至设备。此外,语音合成设备10可为笔记本式或平板式个人计算机而不是诸如便携式电话的终端设备。

[0108] 如稍后将详细描述,语音输入部102包括:麦克风,其将用户输入的语音转换为电语音信号;LPF(低通滤波器),其将被转换的语音信号的高频成分除去;以及A/D转换器,其将除去了高频成分的语音信号转换为数字信号。因此,语音输入部102被构造为接收用户讲出的询问或话语的输入语音信号。说话区间检测部104通过对被转换为数字信号的语音信号进行处理来检测说话区间。

[0109] 音高分析部106对被检测为说话区间的语音信号执行频率分析。通过所述频率分析而获得的第一共振峰中,音高分析部106获得所述语音信号的特定区间(第一区间)的音高并输出指示所获得的音高的音高数据。注意,这里的第一区间是例如询问(或话语)的词尾,并且所述第一共振峰是例如通过对语音进行频率分析而获得的多个共振峰中的一个,其频率最低;在图2示出的示例中,第一共振峰是其结尾用参考标记“A”指示的峰区。对于频率分析,可以采用FFT(快速傅里叶变换)或任何其他期望的常规已知方法。稍后将详细描述对询问(或话语)的词尾进行识别的具体技术。

[0110] 语言分析部108参考在语言数据库122中预先存储的音素模型来确定被检测为说话区间的被检测语音信号接近哪个音素,从而分析(识别)语音信号所定义的词汇的含义。注意,隐马尔科夫模型可以用作音素模型。

[0111] 答复创建部110参考答复数据库124和信息获取部126来创建与语言分析部108识别出的含义相对应的答复。例如,响应于询问(或话语)“现在几点?”,语音合成设备10从内置的实时时钟(未示出)获取时间信息并且从答复数据库124获取不同于时间信息的其他信息。以这种方式,答复创建部110可以创建答复(响应)“现在是 xx点xx分”。

[0112] 另一方面,响应于询问“ashita no tenki wa”(日语罗马字,意思是“明天天气如何?”),语音合成设备10无法创建答复,除非其访问外部服务器来获取天气信息。即,在语音合成设备10无法创建答复的情况下,信息获取部126被构造为或配置为经由因特网访问外部服务器以获取创建答复所需的信息。即,答复创建部110被构造为或配置为从答复数据库124或外部服务器获取针对询问的答复(响应)。以前述方式,答复创建部110、答复数据库124、信息获取部126等的组合充当被构造为获取针对话语的答复的获取部。

[0113] 在本实施例中,答复创建部110将答复输出为音素列,该音素列为定义了各单独音素的音高和产生时机的语音序列。通过语音合成部112根据这种定义了各单独音素的音高和产生时机的语音序列来合成语音,可以输出答复的基本语音。然而,在本实施例中,由语音序列定义的基本语音在语音合成部112对该基本语音进行改变之后才输出。

[0114] 语音合成部112通过以如下方式改变整个语音序列的音高来执行语音合成:使由答复创建部110创建的答复语音序列的特定区间(第二区间)的音高与从音高分析部106供应的第一区间的音高数据呈现出预定音程关系,并且随后语音合成部112输出合成的语音作为答复语音信号。注意,虽然描述了第二区间构成答复的词尾的区间,但是如稍后将描述的,其不限于这种答复的词尾。此外,虽然与第一区间的音高数据具有预定音程关系的音高被描述为从该音高数据表示的音高向下(低)五度的音高,但是其也可以是具有除了从所述音高数据表示的音高向下五度以外的关系的音高。在任何情况下,语音合成部112都充当语音产生部,其不仅被构造为产生所获取的答复的语音,而且被构造为控制答复语音的音高具有与所分析的第一区间的音高相关联的音高。

[0115] 在合成语音时,语音合成部112使用在语音库128中登记的语音片段数据。语音库

128是语音片段数据预先登记在其中的数据库,所述语音片段数据定义了将用作语音素材的各种语音片段(诸如各单独音素以及音素至音素的过渡部分)的波形。更具体地,语音合成部 112按如下方式产生语音信号:将语音序列的各单独声音(音素)的语音片段数据进行组合、对声音(音素)之间的连接部分进行修改并改变答复的词尾的音高。注意,由此通过语音合成部112产生的语音信号被未示出的D/A转换部转换为模拟信号并随后通过扬声器142可听地输出。

[0116] 下面参考图4描述语音合成设备10的行为,图4是示出通过语音合成设备10执行的处理的流程图。首先,一旦用户执行了预定的操作,例如,一旦用户在主菜单上选择了与对话处理相对应的图标等,CPU就激活与该处理相对应的应用程序。通过激活所述应用程序,CPU 建立图1中示出的框图中所示的各部的功能。

[0117] 首先,在步骤Sa11,用户以发声的形式将询问或话语输入至语音输入部102,并且语音输入部102接收输入语音信号。随后,在步骤Sa12,说话区间检测部104将输入语音信号中语音的零强度状态(即,零音量状态)持续超过预定时间段的区间检测为静音区间而将输入语音信号的其他区间检测为说话区间,并且随后说话区间检测部 104将说话区间的语音信号供应至音高分析部106和语言分析部108。

[0118] 随后,在步骤Sa13,音高分析部106对询问(或话语)的检测到的说话区间中的语音信号进行分析,识别询问(或话语)中的第一区间(词尾)的音高,并将随后将指示识别出的音高的音高数据供应至语音合成部112。下面描述凭借音高分析部106对询问(或话语)中的词尾进行识别的一种具体方法的示例。

[0119] 假设对话中给出询问(或话语)的人想要得到针对该询问(或话语)的答复,则认为询问(或话语)中构成词尾的一部分与其他部分相比音量将临时变大。因此,可以例如用如下方式识别出第一区间(词尾)的音高。

[0120] 首先,音高分析部106将被检测为说话区间的询问(或话语)的语音信号转换为分别针对询问(或话语)的音量和音高的波形。图 5的(a)示出了语音信号的音量波形的示例,其中纵轴表示语音信号的音量而横轴表示经过的时间。图5的(b)示出了语音信号的音高波形的示例,其中纵轴表示通过对同一语音信号进行频率分析而获得的第一共振峰的音高而横轴表示经过的时间。注意,图5的(a)的音量波形和图5的(b)的音高波形的时间轴相同。如所示的,音量波形具有多个局部最大值(局部峰)。

[0121] 第二,音高分析部106识别图5的(a)的音量波形的时间上最后的局部峰P1的时间。

[0122] 第三,音高分析部106将所识别出的局部峰P1的时间前后的预定时间范围(例如,100微秒至300微秒)识别为词尾。

[0123] 第四,音高分析部106从图5的(b)所示的音高波形中提取并输出与识别出的词尾相对应的区间Q1的平均音高作为第一区间的音高数据。

[0124] 据认为如上将说话区间中音量波形的最后局部峰P1识别为与询问(或话语)的词尾相对应的时间能够降低对对话中的询问的词尾的错误检测。

[0125] 虽然已将图5的(a)中示出的音量波形中的时间上最后的局部峰P1前后的预定时间范围描述作为词尾,但是将局部峰P1作为其开始时间点或结束时间点的预定时间范围可以被识别为词尾。此外,可以提取并输出区间Q1的开始时间点或结束时间点处的音高或者局部峰P1处的音高作为第一区间的音高数据,而不是将与识别出的词尾相对应的区间Q1的

平均音高作为所述音高数据。

[0126] 同时,在步骤Sa14,语言分析部108对输入语音信号中的词汇的含义进行分析,并向答复创建部110供应指示词汇的含义的数据。随后,在步骤Sa15,答复创建部110通过利用答复数据库124创建与所分析的词汇含义相对应的适当的答复词汇或者必要时经由信息获取部126从外部服务器获取该适当的答复词汇来获取答复语言数据列,随后基于所获取的语言数据列来创建语音序列(答复语音序列),并且之后将所创建的语音序列供应至语音合成部112。

[0127] 图6的(a)是用五线谱示出了响应于询问(话语)“ashita no tenki wa”(日语罗马字,意思是“明天天气如何?”)而创建的日语答复语音序列中的示例基本音高变化的图。在图6的(a)的示出示例中,在答复“hare desu”(日语罗马字,意思是“明天是晴天”)中为每个声音(音节)分配了音符,从而指示由语音序列定义的基本语音的各单独词汇(音节或音素)的音高和产生时机。虽然为了简化说明在图6的(a)的示出示例中为每个声音(音节或音素)分配了一个音符,但是可以利用音符连接符(比如连音符或滑音)为给定的一个声音分配多个音符,可以将不同音高的多个声音瞬态互连等(例如参见图6的(c))。

[0128] 接下来,在步骤Sa16,语音合成部112在从答复创建部110供应的答复语音序列中识别预定的第二区间(在本示例中为说话的词尾)的音高(初始音高)。例如,在图6的(a)中,参考标记“B”示出的音符指示了在答复“hare desu”中的词尾(第二区间)处的区间“su”(日语罗马字)的音高(初始音高);在图6的(a)的示出示例中,第二区间的音高为“F”。可以用任何适当方案来执行从所获取的答复语音序列中提取第二区间(在示出示例中的词尾)。例如,由于答复语音序列中的各单独音节可以彼此区分开,因此最后音节的区间可以提取作为第二区间(词尾)。

[0129] 随后,在步骤Sa17,语音合成部112以如下方式改变语音序列中所定义的音高:使语音序列中所定义的词尾的初始音高呈现出从供应自音高分析部106的音高数据所指示的音高向下(低)五度的音程关系。

[0130] 图6的(b)是以五线谱示出了答复语音“hare desu”(日语罗马字)的示例,该答复语音的音高从图6的(a)所示的音高向下(低)移动了五度。在图6的(b)中,由参考标记“A”示出的音符指示了询问“ashita no tenki wa?”(日语罗马字)的词尾处的“wa”(日语罗马字)区间(第一区间)的示例音高;在该示出示例中,询问的词尾的音高为“G”。在这种情况下,语音合成部112以如下方式移动整个语音序列中的各音高:将答复“hare desu”(日语罗马字)中的由参考标记“B”示出的词尾处的“su”区间的音高改变为从音高“G”向下(低)五度的“C”。

[0131] 随后,在步骤Sa18,语音合成部112对由此改变后的语音序列的语音进行合成并输出合成的语音。虽然没有具体示出,但是一旦答复的语音(答复语音)被输出,则CPU终止当前应用程序的执行并返回至主菜单。

[0132] 答复语音序列的音高变化模式(音符模式)不限于图6的(a)和图6的(b)中示出的,并且可以依照要求设置任何其他音高变化模式。图6的(c)和图6的(d)示出了以不同于图6的(a)和图6的(b)示出的音高变化模式的音高变化模式发出同一语言数据列“hare desu”(日语罗马字)的声音的示例。

[0133] 图6的(a)至图6的(d)示出了日语的示例答复。显而易见的是,即使答复的含义是

相同的,如果用于答复的语言种类不同,则语音完全不同,因此,包括语调和重音在内的音高变化模式(音符模式)也不同。因此,下面将讨论针对多个不同种类语言的示例答复语音。图6的(e)以五线谱示出了用英语做出答复“hare desu”(即,“It’ll be fine.”)的情况下经过了音高移动的示例音高变化模式(音符模式)。图6的(f)示出了在用中文做出答复“hare desu”(即,“是晴天”)的情况下经过了音高移动的示例音高变化模式(音符模式)。

[0134] 图7是对语音合成设备10的实施例给予用户的各种印象进行说明的图。如图7的(a)所示,用户W将询问“ashita no tenki wa?”(日语罗马字)输入至作为终端设备的语音合成设备10。如果构成该询问的词尾的“wa”的音高为“G”,则在本实施例中构成针对该询问的答复的语音序列“hare desu”(日语罗马字)的语音在音高移动的情况下被合成,使得答复的词尾处的“su”(日语罗马字)的音高变为“C”,如图7的(c)所示。以这种方式,本实施例可以给予用户W好似正在实际进行对话的良好印象,而不会给予用户不自然感觉。

[0135] 另一方面,如果语音序列“hare desu”(日语罗马字)的语音在音高没有移动(参见图6的(a))的情况下被合成,则答复的词尾处的“su”(日语罗马字)被输出为具有音高“F”,如图7的(c)所示。在这种情况下,音高“F”与构成询问“ashita no tenki wa?”(日语罗马字)的词尾的“wa”(日语罗马字)的音高“G”为非协和音程关系。即,参见图3,音高“G”的频率(396.0Hz)与“F”的频率(352.0Hz)处于9/8关系。因此,将给予用户W不好的印象(比如厌恶),而不是不自然的感觉。但是,注意,语音合成设备10可以被构造为明确地给予用户这种不好印象。

[0136] <第二实施例>

[0137] 下面描述本发明的语音合成设备10的第二实施例,其采用了所述答复语音产生方法的修改形式。图8是示出本发明的语音合成设备10的第二实施例的构造的框图。虽然上述第一实施例以如下方式被构造,即,答复创建部110输出其中对响应于询问的答复语言的每个声音(音节)分配音高的语音序列并且语音合成部112基于该语音序列对答复的语音(答复语音)进行合成,但是第二实施例以如下方式来构造,即,答复语音输出部113获取针对询问(话语)的答复(响应)并且产生和输出整个答复(响应)的语音波形数据。

[0138] 上述答复(响应)的示例包括由答复语音输出部113所创建的答复、从外部服务器所获取的答复、从提前准备的多个答复之中所选择的答复。此外,上述语音波形数据例如是WAV格式的数据,其中各声音(音节)没有像前述语音序列中的划分那样与其他声音(音节)划分开;即,没有按音节明确地分配音高。图9的(a)示出了这种答复语音波形数据的示例音高变化模式,其对应于如图6的(a)的基本语音。因此,如果将包括这种基本语音的语音波形数据简单地再现,则再现的语音波形数据不会恰当地对应于询问(话语或讲话输入)的音高变化,并且将给出机械感觉。鉴于此,第二实施例被构造为与第一实施例一样根据询问(话语或讲话输入)的音高变化来对语音波形数据的音高进行控制。为此,在第二实施例中设置后处理部114。

[0139] 后处理部114以如下方式执行控制以改变答复语音的整个语音波形数据的音高:使答复(响应输出)的词尾的音高与询问(讲话输入)的词尾的音高呈现出预定的音程关系(例如,协和音程关系)。更具体地,后处理部114分析从答复语音输出部113输出的答复语音波形数据(基本语音)中的词尾(第二区间)处的音高(初始音高),随后后处理部114以如下方式对从答复语音输出部113输出的整个语音波形数据(基本语音)执行音高转换:使词尾

(第二区间)的音高与从音高分析部106供应的输入语音的词尾(第二区间)的音高呈现出预定的音程关系(例如,向下(低)五度)。即,后处理部114 将所获取的答复(响应)的词尾的音高改变为从询问(话语)的词尾的音高向下五度,其是与询问(话语)的词尾的音高的示例协和音程关系。如图9的(b)所示,音高转换的结果实质上类似于图6的(b)所示的音高移动的结果。注意,为了对答复语音波形数据中的词尾(第二区间)处的音高(初始音高)进行分析,后处理部114可以使用与以上参照图5描述的方法类似的方法。即,后处理部114可以对答复语音波形数据的音量进行分析以检测最后的局部最大值(局部峰)并将包含最后的局部最大值的适当时间范围识别为词尾(第二区间)。

[0140] 由于如上构造的第二实施例不包括用于执行复杂语音合成的机构(比如语音合成部112),因此其非常适合应用至答复语音的词汇不复杂的情况(例如,包括简单非正式反馈的答复)。即,在针对询问的答复不需要太复杂或明确的情况下,例如,用简单答复(比如“hai”(是)或“iie”(不))或非正式反馈(比如“sou desune”(“我同意。”))对询问进行答复的情况下,答复语音输出部113 只需要从多个预先存储的语音波形数据中选择和输出适于询问的语音波形数据,因此,可以简化语音合成设备的构造。

[0141] 在第二实施例中,答复语音输出部113和后处理部114的组合充当了语音产生部,其不仅被构造为产生所获取的答复的语音,而且还被构造为以如下方式对所获取的答复的语音的各音高进行控制:使该语音的第二区间具有与所分析的第一区间的音高相关联的音高。

[0142] <应用示例和修改例>

[0143] 应当认识到,本发明不限于上述第一和第二实施例并且本发明的以下各种其他应用示例和修改例也是可行的。此外,可以酌情组合多个应用示例和修改例中的任意选择的应用示例和修改例。

[0144] <语音输入部>

[0145] 虽然已经关于语音输入部102经由麦克风输入用户的语音(话语)并且将输入语音(话语)转换为语音信号的情况对本发明的实施例进行了描述,但是本发明不限于此,语音输入部102可以被构造为经由记录介质、通信网络等接收被另一处理部处理的或从另一装置供应(或转发)的语音信号。即,可以按照任何期望方式对语音输入部 102进行构造,只要其接收询问或话语的输入语音信号即可。

[0146] <答复等中的词首或词尾>

[0147] 虽然描述了第一和第二实施例被构造来对与询问的词尾的音高相对应的答复(响应)的词尾的音高进行控制,但是取决于语言、方言、词汇、说话方式等,答复(响应)中不同于词尾的另一部分(诸如词首)有时可以变为有特点或特色的。在这种情况下,当给出询问的人接收到针对该询问的答复(响应)时,通过无意识地对询问的词尾的音高和答复(响应)的有特点的词首的音高进行比较来判断关于该答复(响应)的印象。因此,在这种情况下,可以进行配置以对与询问的词尾的音高相对应的答复(响应)的词首的音高进行控制。如果答复(响应)的词首是有特点的,则这种配置可以给接收该答复(响应)的用户一种心理印象。

[0148] 同样原理可以用于答复或话语,并且可预期的是印象有时是基于词首以及词尾处的音高来判断的。对于答复或话语,还可以预期的是印象是基于平均音高、词汇被最大强度地说出的部分的音高等来判断的。因此,可以说询问的第一区间和答复(响应)的第二区间

不必限于词尾和词首。

[0149] <音程关系>

[0150] 虽然本发明的上述实施例被构造为对语音合成进行控制以使得答复(响应)的词尾等的音高变为从询问的词尾等的音高向下五度,但是本发明可以被构造为对答复(响应)的词尾等的音高进行控制以呈现不同于从询问的词尾等的音高向下五度的协和音程。例如,不同于从询问的词尾等向下五度的协和音程可以是完全八度(八度)、完全五度、完全四度、大三度或小三度、或者大六度或小六度。

[0151] 此外,可以凭经验辨识一些即使与询问的词尾等的音高不是协和音程关系但仍可以给出良好(或不好)印象的音程关系。因此,本发明可以构造为对答复的音高进行控制以呈现出任一这种音程关系。然而,同样在这种情况下,当询问的词尾等的音高和答复(话语)的词尾等的音高彼此相距甚远时,针对该询问的答复易于变得不自然,因此,期望的是答复的音高在从询问的音高向上一个八度和向下一个八度的范围内。

[0152] <答复的音高移动>

[0153] 在上述构造中,控制由语音序列或语音波形数据所定义的答复(响应)的词尾等的音高从而与询问的词尾等的音高呈现出预定的关系,更具体地,如在上述实施例中,将答复的词尾等的音高改变为例如从询问的词尾等的音高向下五度,因此,可能不期望地以不自然地低的音高合成答复的语音。下面描述用于避免这种不便的应用示例(应用示例1和应用示例2)。

[0154] 图10是示出应用示例1中的处理的主要部分的流程图。应用示例1中的处理的主要部分对应于图4中示出的步骤Sa17处的“答复的音高确定”中所执行的操作。即,在应用示例1中,图10示出的处理在图4中示出的步骤Sa17处被执行,如稍后将详细描述。

[0155] 首先,在步骤Sb171,语音合成部112获得并临时性确定答复的音高(答复音高),其例如是从供应自音高分析部106的音高数据所指示的音高向下五度的音高。然后,在步骤Sb172,语音合成部112确定所述临时性确定的音高是否低于预定的阈音高(第一阈音高)。注意,该阈音高(第一阈音高)被设置为例如下与要在语音的合成中使用的下限频率相对应的音高,若低于其则会给出不自然感觉的音高。

[0156] 如果临时性确定的音高(即,从答复的词尾的音高向下五度的音高)低于预定的阈音高(即,步骤Sb172处“是”判定),则在步骤Sb173处语音合成部112将临时性确定的音高移动至从该临时性确定的音高向上(高)一个八度的音高。另一方面,如果临时性确定的音高等于或高于预定的阈音高(即,步骤Sb172处“否”判定),则跳过步骤Sb173的操作。

[0157] 随后,在步骤Sb174,语音合成部112正式确定答复的音高将要移动至的目标音高作为答复的音高。即,如果临时性确定的音高低于阈音高,则语音合成部112正式确定临时性确定的音高向上移动一个八度的音高作为目标音高。如果临时性确定的音高等于或高于阈音高,则语音合成部112正式确定临时性确定的音高直接作为目标音高。在步骤Sb174之后,处理回到图4的步骤Sa18。由此,语音合成部112对包括正式确定的音高的语音序列的语音进行合成并输出由此合成的语音。

[0158] 根据该应用示例1,在要改变的音高低于阈音高的情况下,将要改变的音高向上移动一个八度,由此,可以避免以不自然地低的音高合成答复的语音的不便。

[0159] 虽然在上面描述了应用示例1将答复的词尾等的音高向上移动一个八度,但是本

发明不限于此。例如,如果临时性确定的音高高于另一预定的阈音高(第二阈音高),则临时性确定的音高可以向下移动一个八度。更具体地,如果用户给出的询问的词尾等的音高过高,则有时从临时性确定的音高向下五度的音高也可能过高。在这种情况下,将以不自然地高的音高合成答复的语音。如果从由音高数据所指示的音高向下五度的音高(临时性确定的音高)高于所述另一预定的阈音高(第二阈音高),则为了避免所述不便,只要将答复的词尾等的音高从临时性确定的音高向下移动一个八度即可满足。可以进行上述将答复的词尾等的音高向上移动一个八度的修改例和上述将答复的词尾等的音高向下移动一个八度的修改例中的至少一个。

[0160] 此外,在某种情况下,语音合成可以输出具有预定性别和年龄(儿童或成人)的虚拟人物的语音的答复。如果在指定了女性或儿童人物的情况下如上所述地将答复的词尾的音高从询问的词尾的音高一律向下移动五度,则答复的语音将以不适合该人物的低音高而合成,因此,可以将答复的词尾的音高向上移动一个八度。

[0161] 图11是示出这种应用示例(应用示例2)中的处理的主要部分的流程图。应用示例2中的处理的主要部分对应于图4示出的步骤Sa17处的“答复的音高确定”中执行的操作。图11示出的应用示例2中的处理不同于图10中示出的应用示例1中的处理的地方在于,语音合成部112在步骤Sb171处获得和临时性确定从供应自音高分析部106的音高数据所指示的音高向下五度的音高,并随后在步骤Sc172处确定当前是否指定了“女性”或“儿童”作为定义询问中的人物的属性。

[0162] 如果“女性”或“儿童”当前被指定为属性(步骤Sc172处“是”判定),则语音合成部112在步骤Sb173将临时性确定的音高向上移动一个八度。另一方面,如果“女性”或“儿童”当前未被指定为属性,并且“男性”或“成人”当前被指定为属性(步骤Sc172处“否”判定),则跳过步骤Sb173的前述操作。后续操作与应用示例1中的相同。

[0163] 根据该应用示例2,如果做出利用女性或儿童语音来产生答复的设置,则将临时性确定的音高向上移动一个八度,由此,可以避免以不自然地低的音高合成答复的语音的不便。

[0164] 虽然在上面描述了应用示例2被构造成在“女性”或“儿童”当前被指定为属性的情况下将临时性确定的音高向上移动一个八度,但是,例如,在“男性”当前被指定为属性的情况下,临时性确定的音高可以向下移动一个八度,以避免将以不适合该人物的音高合成答复的语音的不便。

[0165] <不协和音程>

[0166] 虽然上述实施例被构造为以如下方式对语音合成进行控制:使针对询问的答复(响应)的词尾等的音高与询问的词尾等的音高呈现出协和音程关系,但是语音合成可以以如下方式进行控制:答复(响应)的词尾等的音高与询问的词尾等的音高呈现出不协和音程关系。如果以与询问的词尾等的音高具有不协和音程关系的音高来合成答复的语音,则可发生给予给出询问的用户不自然感觉、不好印象、厌恶感觉等使得无法建立流畅的对话的不良的可能性。但是,反之,也存在认为这些感觉有助于压力释放的观点。因此,在本发明中,可以准备期望良好印象等的答复的操作模式(第一模式)和期望不好印象等的答复的另一操作模式(第二模式),使得根据第一和第二模式中的任一个来对语音合成进行控制。

[0167] 此外,图12是示出该应用示例(应用示例3)中的处理的主要部分的流程图。应用示

例3中的处理的主要部分对应于图4示出的步骤Sa17处的“答复的音高确定”中所执行的操作。图12中示出的应用示例3中的处理与图10中示出的应用示例1中的处理的不同之处如下。即,应用示例3中的语音合成部112在步骤Sd172处确定当前是否设置了第一模式作为操作模式。

[0168] 如果当前设置了第一模式作为操作模式(步骤Sd172处“是”判定),则语音合成部112在步骤Sd173A将答复的例如词尾的音高设置为与询问的例如词尾的音高处于协和音程关系。另一方面,如果当前设置了第二模式作为操作模式(步骤Sd172处“否”判定),则语音合成部112在步骤Sd173B将答复的词尾的音高设置为与询问的词尾的音高处于不协和音程关系。后续操作与应用示例1和应用示例2中的相同。

[0169] 即,根据应用示例3,如果当前设置了第一模式,则以与询问的词尾处于协和音程关系的音高来合成答复(响应)的语音,而如果当前设置了第二模式,则以与询问的词尾处于不协和音程关系的音高来合成答复(响应)的语音。由此,根据应用示例3,用户可以按情况适当使用两种模式中的任一种。注意,可以按任何期望方式来执行第一模式和第二模式中任一种的设置,例如,通过用户进行的选择操作或者通过基于例如在应用程序的执行期间在设备内产生的指令的自动选择。

[0170] 虽然应用示例1、应用示例2和应用示例3在上面已经关于使用与第一实施例中使用的语音序列相类似的语音序列的情况进行了描述,但是显然这些示例可以使用与第二实施例中使用的语音序列相类似的语音序列。

[0171] <答复的语音>

[0172] 虽然上述实施例被构造为通过人类语音的合成来获得答复的语音(即,以合成的人类语音来产生答复),但是也可以通过动物的语音的合成来获得答复的语音(即,以合成的动物语音来产生答复)。即,本文使用的术语“语音”指的是包含动物的语音以及人类的语音的概念。因此,下面描述应用示例4,其中通过动物的语音的合成来获得答复的语音(即,以合成的动物语音来产生答复)。

[0173] 图13是概括应用示例4中执行的处理的图。在通过动物的语音的合成来获得答复的语音的情况下,该处理被构造为响应于询问的词尾的音高而仅将动物的语音的词尾的音高移动至预定的音高。因此,不需要执行对询问的含义进行分析的操作、获取与所分析的含义相对应的信息的操作、创建与所述信息相对应的答复的操作等。

[0174] 一旦用户W讲出并输入询问“ii tenki dane?”(日语罗马字,意思是“天气不错,不是吗?”)至语音合成设备10中,则语音合成设备10对构成该询问的词尾的“ne”(日语罗马字)的音高进行分析。如果“ne”的音高为“G”,则语音合成设备10对狗的语音“wan”(日语罗马字)的语音波形数据进行后处理并且将构成“wan”的词尾的“n”的音高改变为“C”,该音高为从询问的词尾的音高向下五度(即,与询问的词尾的音高处于协和音程关系的示例)的音高,从而语音合成设备10输出由此改变的音高“C”。

[0175] 在通过动物的语音的合成来获得答复的语音的情况下,无法从答复中获得用户期望的信息。即,当用户问出了询问“asu no tenki wa?”(日语罗马字,意思是“明天天气如何?”)时,用户无法得到明天的天气信息。然而,如果响应于某个用户的询问以如下方式来合成动物的语音:动物的语音的词尾的音高呈现出例如从询问的词尾的音高向下五度的音程关系,则动物的语音可以给用户舒服、安心的良好印象;就这一点而言,该应用示例与以

合成的人类语音产生答复的情况相同。因此,即使以动物的语音产生答复,也可以预期到给用户好似该用户正实际与发出该语音的虚拟动物进行沟通的某种治疗效果。

[0176] 作为示例,可以在语音合成设备10上设置显示部使得可以在该显示部上显示虚拟动物,如图13的(b)所示。同时,可以以摇动尾巴、倾斜头部等的动画图像的形式在所述显示部上显示虚拟动物。这种配置可以进一步强化所述治疗效果。

[0177] 如果其语音要被合成的动物是狗,则该应用示例可以被构造为允许从各种狗类型(诸如吉娃娃、博美犬、金毛猎犬之类)之中进行选择。如上所述的被构造为用合成的动物的语音创建答复的语音合成设备10可以应用至不同于终端设备的模仿期望动物的诸如机器宠物、填充玩具等。

[0178] <其他>

[0179] 在上述实施例中,在语音合成设备10中设置了被构造为获取针对询问(话语)的答复(响应)的语言分析部108、语言数据库202 和答复数据库204。然而,鉴于终端设备中的处理负荷变大和存储能力受限等不便,上述语言分析部108、语言数据库202和答复数据库204可以设置在外部服务器中。即,只需要将语音合成设备10中的答复创建部110(答复语音输出部113)构造为以某种方式获取针对询问的答复并输出该答复的语音序列(语音波形数据),而答复是在语音合成设备10中还是在不同于语音合成设备10的结构(诸如外部服务器)中创建则完全不重要。在可以在语音合成设备10中创建针对询问(话语)的答复(响应)而不访问外部服务器等的应用中,可以无需信息获取部126。

[0180] <第三实施例>

[0181] 下面,将参照图14描述本发明的第三实施例。本发明的第三实施例的特征在于根据包括在询问(话语)的输入语音信号中的非语言(非言语)信息来控制语音合成。图14是示出本发明的语音合成设备10的第三实施例的构造的框图。用与图1中相同的参考标记示出与图1中的块实现基本相同功能的图14中的块,并且对其不进行描述以避免不必要地重复。除图1示出的构造外,图14的第三实施例还包括非语言分析部107和语音控制部109。

[0182] 非语言分析部107对被说话区间检测部104检测为说话区间的语音信号进行分析并输出不同于音高(不同于音高信息)的非语言信息。虽然语音的速度(说话速度)、质量和音量、给出询问的用户的面部表情、姿势和衣着等可被假定为非语言信息的示例,但是在本实施例中假定语音的速度和音量作为非语言信息进行分析。即,在本实施例中非语言分析部107对询问(话语)的速度和音量进行分析并输出所分析的速度和音量作为不同于音高信息的非语言信息。注意,非语言分析部107所分析的输入语音信号的音量被供应至音高分析部106(如虚线所指),从而该音量可以被用于识别第一区间的音高的音量信息。

[0183] 语音控制部109根据从音高分析部106输出的音高数据和从非语言分析部107输出的非语言信息(速度和音量)来控制语音合成部112。下面参照图15描述语音控制部109如何控制语音合成部112。

[0184] 图15是语音合成设备10的第三实施例中执行的处理的流程图。图15被示出为图4的修改例,并且执行与图4中的实质相同的操作的步骤用相同参考标记示出并将不在此处进行描述以避免不必要的重复。

[0185] 一旦在步骤Sa12检测到说话区间,则非语言分析部107在步骤Sa19根据检测到的说话区间的语音信号对询问的速度(说话速度)进行分析并输出指示该速度的速度数据。

与步骤Sa19的速度分析相并行,非语言分析部107在步骤Sa20执行下面的音量分析。即,非语言分析部107获得例如检测到的说话区间中的询问(话语)的语音信号的平均音量,从而输出指示该平均音量的音量数据,并且除该音量数据输出外,非语言分析部107还获得语音信号的音量波形。这里获得的音量波形的一个示例类似于图5的(a)示出的音量波形。

[0186] 在步骤Sa17对答复语音的词尾(第二区间)的音高进行改变之后,语音控制部109前进至步骤Sa21。在步骤Sa21,语音控制部109 确定语音合成部112应当合成具有与指示答复(响应)的平均音量的音量数据相对应的音量的整个答复(响应)的语音。随后,在步骤Sa22,语音控制部109确定语音合成部112应当合成与速度数据相对应的速度的整个答复(响应)的语音。此外,在下一步骤Sa18,语音合成部112合成并输出具有所确定的音量和所确定的速度的语音序列的语音,该语音序列的改变已由语音控制部109确定。

[0187] <第四实施例>

[0188] 接下来,将描述作为第三实施例中采用的答复语音产生方法的修改例的本发明的第四实施例。图16是示出本发明的语音合成设备 10的第四实施例的构造的框图。类似于图8示出的语音合成设备10,图16示出的语音合成设备10以如下方式构造:使答复语音输出部113获取针对询问(话语)的答复(响应)并产生整个答复(响应)的语音波形数据,并且后处理部114a执行用于对答复语音的词尾(第二区间)的音高进行改变的控制。用与图8中相同的参考标记示出与图8中的块实现基本相同功能的图18中的块,并且不对其进行描述以避免不必要地重复。图18的第四实施例除了包括图8中示出的构造以外还包括非语言分析部107,并且第四实施例中的后处理部114a 与图8中示出的后处理部114略微不同。图18中的非语言分析部107 被构造为与图14中的非语言分析部107相似。

[0189] 除了以与图8的后处理部114的方式类似的方式(即,使答复(响应输出)的词尾的音高与询问(讲出的输入)的词尾的音高呈现出预定的音程关系(比如协和音程关系))执行用于改变答复语音的整个语音波形数据的音高的控制之外,在第四实施例中的后处理部114a还执行用于再现从答复语音输出部113输出的、具有与所分析的非语言信息的音量相对应的强度和/或具有与非语言信息的速度相对应的速度的整个答复语音波形数据的控制。即,在第四实施例中,后处理部114a除了对答复语音进行音高控制之外还对整个答复语音的音量和再现速度进行控制。

[0190] <第五实施例>

[0191] 在人与人之间的对话中,快速讲出询问(话语)的人通常没有耐心、性急或其他类似的性格或心理状态,因此,认为这个人希望快速获得针对该询问(话语)的答复(响应)。另一方面,缓慢地讲出询问(话语)的人通常例如处于小心冷静状态,因此,认为这个人准备好耐心地等待针对该询问的答复(响应)。

[0192] 因此,第五实施例被构造为根据询问(话语)的非语言信息(即,本例子中的询问(话语)的速度)来控制语音合成开始时机(即,从询问的结尾到开始对答复进行语音合成的时间间隔或停顿)。

[0193] 语音合成设备10的第五实施例可以大致与图14中示出的第三实施例相同,只要其被构造为使用语音序列即可。但是,第五实施例中的语音控制部109对语音合成部112进行控制以随着由非语言信息的速度数据所指示的询问的速度的降低而延迟语音合成开始时机。

[0194] 图17是对第五实施例中执行的语音合成进行说明的图。如果用户W快速输入询问“ashita no tenki wa?”(日语罗马字,意思是“明天天气如何?”)至语音合成设备10,则将在相对早的时机以语音输出示例答复“hare desu”(日语罗马字,意思是“明天是晴天。”),并且答复的整体速度将是快的,如图17的(a)所示。另一方面,如果用户W缓慢地输入相同询问“ashita no tenki wa?”至语音合成设备10,则将在相对晚的时机以语音输出答复“hare desu”,并且答复的整体速度将是缓慢的,如图17的(b)所示。

[0195] 注意,在图中,从询问被讲出的时间到开始答复的语音合成的时间的时间间隔或停顿是由从询问的词尾到答复的词首的时间段 T_a 或 T_b ($T_a < T_b$) 来表示的。显然,该停顿可以以任何其他适当方式而不是上述方式来进行定义。在图中,询问中以及答复中的文字(字母)之间的较小间隔表示该询问和答复被快速讲出。

[0196] 通过以上方式构造的第五实施例,可以给想要得到针对询问的答复的用户具有自然感觉和符合用户的性格和心理状况的好似在人与人之间进行对话的答复。

[0197] 如果使用语音波形数据代替语音序列,则以与图16中示出的第四实施例实质上相同的方式构造第五实施例。在这种构造中,后处理部114a随着非语言信息的速度的降低而对语音输出的开始时机进行延迟。此外,虽然第五实施例被构造为根据询问的速度来控制语音输出的开始时机,但是对语音输出的开始时机进行控制所依据的速度可以被替换为音量或音高,并且可以酌情结合除速度、音量和音高以外的其他信息。

[0198] <第六实施例>

[0199] 与第一实施例和第二实施例的上述各示例和/或修改例类似的应用示例和/或修改例可以应用至第三、第四和第五实施例。例如,图18是示出作为本发明第六实施例的在应用与图12示出的操作模式类似的操作模式的情况下的答复音高确定程序(routine)的一个示例的流程图。图18类似于图12,除了在图18中增加了步骤Sd174A 和Sd174B之外。

[0200] 图18示出的示例被构造为在考虑了针对询问的答复的“停顿”的情况下执行控制。通常,在人与人之间的对话中,如果一个人试图以故意给出不好印象的方式来对询问进行答复,则对询问的答复的停顿与其他情况相比会大大缩短或大大延长。在比如争论场景中,例如,一个人趋向于快速地(实质上没有停顿地)或故意长停顿地反驳另一人的话语。在图18示出的所示示例中,在考虑这种“停顿”的情况下执行控制。

[0201] 如果在图18的示例中当前设置第一模式作为操作模式,则语音控制部109在步骤Sd173A之后前进至步骤Sd174A,在该步骤中其确定语音合成应当在询问(话语)的词尾之后的预定时间范围内开始,作为要给予语音合成部112的指令。因为“预定时间范围”处于0.5至2.0秒的范围中,因此只要前述停顿在该范围内就认为其是恰当的停顿。

[0202] 另一方面,如果当前第二模式被设置作为操作模式(即,在步骤Sd172处“否”判定),则语音控制部109在步骤Sd173B确定答复的词尾的音高应当与询问的词尾的音高呈现出不协和音程关系。随后,在步骤Sd174B,语音控制部109确定语音合成应当在询问的词尾之后的预定时间范围以外开始,作为要给予语音合成部112的指令。

[0203] 在步骤Sd174A或Sd174B之后,处理返回至图15的步骤Sa21,从而语音控制部109在步骤Sa22确定答复的整体音量。此后,在步骤Sa18,语音合成部112以确定的时机和速度和确定的音量合成其改变已由语音控制部109确定的语音序列的语音,并随后将由此合成的语音输出。

[0204] 即,根据图18中示出的第六实施例,当操作模式为第一模式时将“停顿”控制为在预定时间范围内,而当操作模式为第二模式时,将“停顿”控制为在预定时间范围以外。由此,用户可以酌情使用这两种模式中的任意期望的一种,从而可以增强所述设备的可用性。图 18中示出的应用示例不仅可应用至第三实施例(被构造为基于语音序列按音节进行答复语音合成)而且还可应用于第四实施例(被构造为产生整个答复语音的答复语音波形数据)。

[0205] <第七实施例>

[0206] 虽然第三至第六实施例中的每一个被描述为使用与用户讲出的询问(话语)自身有关的信息(诸如音高、音量和速度)作为非语言信息,但是其他信息(诸如面部表情、姿势和衣着)可以被用作非语言信息。因此,下面描述使用非语音信息作为非语言信息的本发明的第七实施例。

[0207] 图19是示出本发明的语音合成设备10的第七实施例的构造的框图,其特征在于包括成像部130,用于对语音合成设备10周围的区域进行成像。在本实施例中,非语言分析部107根据成像部130 获取的静态图像来分析用户的面部表情(笑、尴尬、生气等)并按如下在语音合成处理中反映所述分析的结果。

[0208] 当用户笑着讲出询问时,例如,音高分析部106易于在对询问进行的音高分析中检测到高音高。由此,非语言分析部107使得音高分析部106将检测到的音高调整为较低音高。另一方面,当用户生气地讲出询问时,音高分析部106易于在对询问进行的音高分析中检测到低音高。因此,非语言分析部107使得音高分析部106将检测到的音高调整为较高音高。通过如此根据用户的面部表情来对音高分析部 106的音高分析的结果进行调整,本实施例可以实现音高检测和分析的更高精确度。

[0209] 此外,语音合成设备10以如下方式构造:如果用户在语音合成设备10已合成和输出答复(响应)的语音之后具有尴尬的面部表情,则将答复的词尾的音高改变为呈现出从询问的词尾的音高向下五度的音程关系。然而,可预期到的是这种音程关系未被正确地保持。因此,非语言分析部107使得语音控制部109将答复与询问的向下五度音程关系改变为另一音程关系,比如从询问的词尾的音高向下四度。作为另一种选择,非语言分析部107使得语音控制部109改变音高分析部106中的音高分析算法(例如,音高波形的哪个局部最大值应当被设置为词尾;应当确定词尾的哪个部分的音高;等等),这是因为还可以预期到的是音高分析部106进行的音高分析的精确度已被降低了。通过如上所述地根据用户的面部表情来对音高分析和语音合成进行控制,本实施例可以将其语音将要被合成的答复导向不会给予不舒服感觉的方向。

[0210] 虽然在上面已关于根据用户的面部表情来对音高分析和语音合成进行控制的情况描述了本实施例,但是本实施例可以构造为根据成像部130获取的用户的图像来分析用户的姿势、衣着等,从而根据该用户的姿势、衣着等来对音高分析和语音合成进行控制。此外,可以设置湿度传感器等替代成像部130或者除了成像部130外还设置湿度传感器等,从而将经由湿度传感器等获取的周围环境的信息用于对语音合成进行控制。例如,在炎热而潮湿的环境中,可以合成预期会缓和沮丧感的答复的语音。

[0211] <应用的规则的波动性等>

[0212] 第三至第六实施例中的每一个被构造为基于询问(话语)的音量和速度并根据上

述规则来确定答复的音量、速度和停顿。然而,在规则固定的情况下,答复易于给予用户该答复单调的印象。因此,可以进行配置以赋予规则波动性,从而在例如所创建的答复不会给出不自然感觉的范围内修改规则,或者可以预先准备该范围内的多个规则,从而可以选择任一规则应用至语音合成。这种配置可以有效避免给用户单调的答复。

[0213] 本发明的第三至第七实施例中的每一个可以总结如下。即,第三至第七实施例中的每一个均包括:非语言分析部107,用于分析不同于与询问(即,话语)有关的音高信息的非语言信息;以及控制部(语音控制部109),用于根据所分析的非语言信息对语音产生部(即,语音合成部112或者答复语音输出部113和后处理部114a的组合)产生答复的语音进行控制。注意,与第一实施例和第二实施例有关的各种应用示例和/或修改例也可以应用于本发明的第三至第七实施例。

[0214] <第八实施例>

[0215] 下面参照图20描述本发明的语音合成设备的第八实施例,其特征在于根据包括在话语(询问)和答复(响应)中的语言信息来控制答复语音的产生。图20是示出本发明的语音合成设备的第八实施例的构造的框图。用与图1中相同的参考标记示出与图1中的块实现基本相同功能的图20中的块,并且不对其进行描述以避免不必要的重复。图20的第八实施例除了包括图1示出的构造以外还包括语音控制部109A,并且第八实施例中的语言分析部108A在构造上不同于图1示出的语言分析部108。

[0216] 第八实施例中的语言分析部108A不仅分析包含在话语(询问)的语音信号中的语言含义,而且分析答复创建部110所创建(获取)的答复(响应)的含义。语言分析部108A将所分析的话语(询问)的含义的结果供应至答复创建部110,并且当从答复创建部110接收到答复(响应)时,语言分析部108A分析所接收到的答复(响应)的含义并且将所分析的该答复(响应)的含义的结果供应至语音控制部109A。

[0217] 在第八实施例中,答复创建部110可以创建(获取)以下类型的答复:

[0218] (1) 指示肯定含义的答复;

[0219] (2) 指示否定含义的答复;

[0220] (3) 针对询问的具体答复;和

[0221] (4) 在具体答复之前发出的共鸣答复(包括非正式答复)。上述类型(1)答复的示例包括“是”或“hai”(日语罗马字),上述类型(2)答复的示例包括“不”或“ie”(日语罗马字)。上述类型(3)答复的示例包括具体地对比如“asu no tenki wa?”(日语罗马字)进行回答的答复。此外,上述类型(4)答复的示例包括“eto”(日语罗马字,意思是“让我想想”)和“sou desune”(日语罗马字,意思是“恩,……”)。答复创建部110所创建的答复语音序列被供应至语音控制部109A和语音合成部112。

[0222] 语音控制部109A根据从音高分析部106供应的音高数据和从语言分析部108A供应的答复的含义来确定要对答复语音序列执行的控制的内容。下面参照图21描述语音控制部109A进行的控制的内容。

[0223] 图21是语音合成设备10的第八实施例中执行的处理的流程图。图21示出为图4的修改例,并且执行与图4中的步骤实质相同的操作的步骤用与图4中相同的参考标记示出并将不进行描述以避免不必要的重复。

[0224] 在图21的步骤Sa14,语言分析部108A对语音信号所定义的词汇(询问或话语)的含

义进行分析并将指示含义的数据供应至答复创建部110。随后,在步骤Sa15a,答复创建部110利用答复数据库124 和信息获取部126 (如有必要) 来创建与所分析的词汇(询问)对应的答复,并将所创建的答复供应至语言分析部108A。注意,(从答复数据库124读出的)所创建的答复是针对询问的属于类型(1)、(2)和(4)中任一个的答复。虽然未在流程图中示出,但是在类型(4)答复之后创建具体答复(即,类型(3)答复)。在步骤Sa15b,答复创建部110输出所创建或所获取的答复的语音序列。

[0225] 图22的(a)示出了类型(1)的肯定答复“hai”(日语罗马字,意思是“是”)的示例语音序列。在图22的(a)的示出示例中,对答复“hai”的每个声音(即,每个音节)分配音符,从而指示出每个词汇(音节或音素)的音高和产生时机。虽然为了简化说明在示出的示例中为每个声音(即,每个音节或音素)分配了一个音符,但是可以利用音符连接符(比如连音符或滑音)为一个声音分配多个音符,可以将不同音高的多个声音瞬态互连等。

[0226] 图23的(a)示出了类型(2)的否定答复“iie”(日语罗马字)的示例语音序列,而图24的(a)示出了类型(4)的答复“eto”(日语罗马字,意思是“让我想想”)的示例语音序列。

[0227] 语言分析部108A在步骤Sa23对从答复创建部110供应的答复的含义进行分析。在本实施例中,语言分析部108A识别该答复是类型(1)、(2)和(4)中的哪一个并且随后将类型识别结果供应至语音控制部109A。

[0228] 语音控制部109A在步骤Sa24确定从答复创建部110输出的整个答复语音序列的音高,使得从答复创建部110输出的答复语音序列的第二区间的音高与从音高分析部106供应的音高数据的第一区间(例如,词尾)的音高呈现出如下音程关系。

[0229] 更具体地,对于类型(1)的肯定答复,语音控制部109A决定对整个答复语音序列的音高进行移动,使得答复的第二区间(词尾)的音高被改变为从话语(询问)的音高数据的第一区间(例如,词尾)的音高向下五度。

[0230] 对于类型(2)的否定答复,语音控制部109A决定对整个答复语音序列的音高进行移动,使得答复的第二区间(词尾)的音高被改变为从话语(询问)的音高数据的第一区间(例如,词尾)的音高向下六度。即,类型(1)的肯定答复的第二区间的音高被设置成高于类型(2)的否定答复的第二区间的音高。

[0231] 对于类型(4)的答复,语音控制部109A决定以如下方式将整个答复语音序列的音高进行移动:将答复的第二区间(词尾)的音高改变为例如从话语(询问)的音高数据所指示的第一区间(例如,词尾)的音高向下五度。

[0232] 更具体地,考虑询问“asu wa hare?”(日语罗马字,意思是“明天是晴天吗?”)的情况,图22的(b)中参考标记A所示出的词尾处的“re”的音高由音高数据指示为“音高E”。在这种情况下,对于类型(1)的肯定答复“hai”(日语罗马字)的语音序列,语音控制部109A决定将整个答复语音序列中定义的音高进行移动,使得图22的(b)中的参考标记B处示出的答复的词尾处的“i”的音高变为“音高A”,其是从询问的词尾的音高“E”向下五度的音高。

[0233] 此外,对于类型(2)的否定答复“iie”(日语罗马字),语音控制部109A决定将整个答复语音序列所定义的音高进行移动,使得图23的(b)中的参考标记C处示出的答复的词尾处的“e”的音高变为“音高G”,其是从询问的词尾的音高“E”向下六度的音高。

[0234] 还考虑询问“asu no tenki wa”(日语罗马字,意思是“明天天气如何?”)的情况,在图24的(b)中的参考标记A处示出的词尾处的“wa”的音高由音高数据指示为“音高E”。在

这种情况下,对于类型(4)的语音序列“êto”(日语罗马字),语音控制部109A 确定语音序列中定义的音高,使得在图24的(b)的参考标记D处示出的词尾处的“to”的音高变为“音高A”,其是从询问的词尾的音高“E”向下五度的音高。

[0235] 现在返回来参照图21,语音控制部109A在步骤Sa25确定答复语音序列的输出时机,即,“停顿”。更具体地,对于类型(1)的肯定答复,语音控制部109A将从询问的词尾到答复的输出的“停顿”设置在0.5至2.0秒的范围内,或者对于类型(2)的否定答复,设置在比针对类型(1)的范围更长的3秒附近。因此,针对类型(1)的肯定答复的答复语音序列将在比针对类型(2)的否定答复的答复语音序列更早的时机输出。注意,对于类型(4)的答复,语音控制部109A将“停顿”设置在与类型(1)的答复相同的0.5至2.0秒的范围内。

[0236] 在步骤Sa18,语音控制部109A控制语音合成部112来将从答复创建部110供应的语音序列移动至所确定的音高并在所确定的时机输出音高移动后的语音序列。根据这种控制,语音合成部112改变了语音序列的音高并以改变后音高合成和输出答复的语音。

[0237] 对于答复语音输出时机,语音控制部109A可以控制语音合成部 112在由实时时钟(未示出)在询问的词尾处(例如,局部最大值P1 (图5)的时间点)开始测量的时间经过预设时间时的时间点处开始输出语音序列。

[0238] 虽然未特别示出,但是语音合成部112被构造为使得如果其已输出类型(4)的答复语音,则随后其输出类型(3)的具体答复语音。以上述方式,应用程序的执行终止,处理返回至主菜单。

[0239] 图25是对响应于用户讲出的询问而输出类型(1)或(2)的答复语音进行说明的图。更具体地,图25示出了作为终端设备的语音合成设备10针对用户W讲出的询问“asu wa hare”(日语罗马字,意思是“明天是晴天吗?”)输出“hai”或“iie”。

[0240] 在图25的示出示例中,以如下方式执行语音合成:答复“hai”或“iie”的词尾的音高呈现出从询问的词尾的音高向下五度或六度的协和音程关系。因此,该答复可以给予好似正在与人进行对话的印象,而不会给予用户不自然的感觉。

[0241] 在图25的示出示例中的类型(1)的答复“hai”的词尾的音高和类型(2)的答复“iie”的词尾的音高相对于询问的词尾的音高呈现出何种关系已经在上面关于图22和图23进行了描述。否定答复“iie”的词尾的音高趋向于低于肯定答复“hai”的词尾的音高。因此,当在本实施例中答复有否定含义时,考虑到不要使另一人或另一方陷入不好情绪中,以低语音输出否定答复。同样从这个角度,本实施例可以给出好似用户正实际与人进行对话的印象。此外,当输出类型(2)的否定答复时,与类型(1)的肯定答复的音量相比,语音控制部109A可使答复的音量适度(即,降低答复的音量)。

[0242] 此外,在该图中的示出示例中,讲出询问时的时间点(例如,词尾处“re”的时间点)用(a)表示,讲出答复“hai”的词首时的时间点用(b)表示,讲出答复“iie”的词首时的时间点用(c)表示。从询问的讲出到答复“hai”被输出时的时间点的时间段T1短于从询问的讲出到答复“iie”被输出时的时间点的时间段T2;即, $T1 < T2$ 。换言之,在语音合成设备10的本实施例中,答复“iie”的输出略微晚于答复“hai”的输出。因此,根据本发明,可以在类似于与人的真实对话中出现的停顿的情况下以合成的语音输出诸如“hai”和“iie”之类的答复。

[0243] 图26是对针对用户讲出的询问输出类型(3)或(4)的答复语音进行说明的图。更具体地,图26示出了语音合成设备10针对用户讲出的询问“asu no tenki wa”(日语罗马字,

意思是“明天天气如何?”)而输出具体答复的情况。在图26的示出示例中,在输出类型(3)的具体答复之前,响应于询问输出类型(4)的答复“êto”。由于需要访问各种数据库和外部服务器中的任一个来创建具体答复,因此有时可能花费相对长的时间来创建答复。如果在具体答复被创建之前在语音合成设备10中持续相当长的静音状态,则会给予用户焦虑感。然而,在本实施例中,在暂时性输出类型(4)的答复的情况下,不会给予用户这种焦虑感。此外,在人与人之间的实际对话中,在答复具体内容之前讲出这种类型(4)的答复是通常做出的行为。由于本实施例被构造为通过仿效这种通常做出的行为来对答复的语音进行合成,因此可以给予用户好似该用户正实际与人进行对话的印象。

[0244] 在该图的示出示例中,讲出询问时的时间点(例如,询问的词尾处“wa”的时间点)用(d)表示,讲出答复“êto”的词首时的时间点用(e)表示,讲出答复“hare desu”的词首时的时间点用(f)表示。优选的是,从询问的讲出直到答复“êto”时的时间点的时间段T3大致等于类型(1)的答复的时间段T1。如果花费长时间来创建比如“hare desu”的具体答复,则可以重复作为类型(4)的答复的同一答复“êto”或另一答复,从而提供增加了时间长度的停顿。

[0245] <利用语音波形数据的修改例>

[0246] 虽然描述了本发明的第八实施例被构造为输出语音序列(其中按每个声音(每个音节)分配音高)作为针对询问的答复,但是第八实施例可以以与上述第二实施例类似的方式进行修改。即,图20的答复创建部110和语音合成部112的组合(即,利用语音合成技术的语音产生部)可以被替换为答复语音输出部113和后处理部114的组合(即,利用波形数据的语音产生部),并且可以从答复语音输出部113输出例如WAV格式的语音波形数据。

[0247] 在这种情况下,后处理部114可以被构造为例如在执行音高转换(比如滤波处理)之后输出(再现)语音波形数据,使得答复的第二区间(例如,词尾)的音高与话语(询问)的音高数据中的第一区间(例如,词尾)的音高呈现出预定的音程关系。此外,可以利用卡拉OK设备领域公知的在不改变语速(说话速度)的情况下对音高进行移动的所谓的音调控制技术来执行该音高转换。

[0248] 注意,与第一实施例和第二实施例有关的各种应用示例和/或修改例也可应用于本发明的第八实施例。

[0249] 简言之,本发明的第八实施例的特征在于包括:语言分析部(语言分析部108A),其分析包含在询问(话语)和答复中的语言信息;以及控制部(语音控制部109A),其控制语音产生部(即,语音合成部112或答复语音输出部113和后处理部114的组合)进行的答复的语音的产生。

[0250] <第九实施例>

[0251] 下面参照图27描述本发明的语音合成设备的第九实施例,其特征在于根据话语(询问)中随时间的音高变化来控制答复(响应)中的音高变化。图27是本发明的语音合成设备10的第九实施例的功能框图。用与图1中相同的参考标记示出与图1中的块实现基本相同功能的图27中的块,并且不对其进行描述以避免不必要的重复。图27的实施例除了包括图1中示出的构造之外还包括非语言分析部107A和语音控制部109B。

[0252] 非语言分析部107A对由说话区间检测部104检测为说话区间的话语的语音信号执行音量分析和频率分析,从而检测话语中随时间的音高变化,并且输出指示话语中的音高

变化的数据作为非语言信息。指示音高变化的数据被供应至语音控制部109B和答复创建部110。

[0253] 答复创建部110利用非语言分析部107A所分析的指示音高变化的数据并参考答复数据库124和信息获取部126来创建与由语言分析部108分析的话语相对应的答复。如第八实施例,第九实施例中的答复创建部110可以创建(获取)以下类型的答复:

[0254] (1) 指示肯定含义的答复;

[0255] (2) 指示否定含义的答复;

[0256] (3) 针对询问的具体内容的答复;和

[0257] (4) 作为针对询问的非正式反馈的答复(参见图22至图24)。

[0258] 由答复创建部110创建/获取的语音序列被供应至语音控制部109B和语音合成部112,如上述第八实施例(参见图20等)。语音控制部109B根据从音高分析部106供应的音高数据和从非语言分析部107A供应的指示话语中的音高变化的数据来确定要对该语音序列执行的控制的内容。

[0259] 图28是在语音合成设备10的第九实施例中执行的语音合成处理的流程图。图28示出为图4的修改例,并且与图4中的步骤执行实质相同的操作的步骤用与图4中相同的参考标记示出并将不对其进行描述以避免不必要的重复。

[0260] 在图28的步骤Sa13a和Sa13b处,以与图4的步骤Sa13处相同的方式,音高分析部106对询问(或话语)的检测到的说话区间中的语音信号进行分析,识别询问(或话语)中的第一区间(词尾)的音高,并随后将指示识别出的音高的音高数据供应至语音合成部112。即,在步骤Sa13a,音高分析部106对检测出的说话区间中的语音信号进行分析以将该信号转换为分别针对询问(或话语)的音量和音高的波形。这种转换后的音量波形和音高波形的示例类似于上面图5中示出的那些。但是,为了便于说明,在图29中示出了类似于图5中示出的内容的转换后的音量波形和音高波形。

[0261] 在步骤Sa26,非语言分析部107A识别如上在步骤Sa13a中由音高分析部106分析的输入语音(话语或询问)的音高波形中的时间上的音高变化状态。现在将参照图29描述用于在步骤Sa26识别这种时间上的音高变化的示例方法。首先,非语言分析部107A从图29的(b)的音高波形中识别在图29的(a)的音量波形的局部最大值P1的时间点之前预定时间 T_s (例如,0.3秒)的时间点P0处的音高 N_0 。随后,非语言分析部107A评估从音高 N_0 到音高 N_1 的音高变化量(即, $N_1 - N_0$)并将该音高变化量作为指示音高变化状态的数据供应至语音控制部109B和答复创建部110。时间 T_s 可以设置为与话语的词首和词尾之间的时间间隔相对应的可变值,而不是设置为预定的固定值。作为一个示例,与话语的词首相对应的时间点可以设置为时间点P0。此外,被非语言分析部107A识别出的指示音高变化状态的数据的内容不限于等于差值 $N_1 - N_0$ 的音高变化量。例如,可以识别话语中的各单独词汇的音高(针对词汇的音高)的变化模式,并且可以将指示由此识别出的变化模式的数据供应至语音控制部109B和答复创建部110作为指示音高变化状态的数据。

[0262] 回过来参考图28,在图28的步骤Sa15a和Sa15b处执行与图4的步骤Sa15或图21的步骤Sa15a及Sa15b类似的操作。然而,在第九实施例中,答复创建部110以如下方式构造:即使对于包括同一文字列的话语(输入语音),只要音高变化状态不同,其就可以根据话语(输入语音)的音高变化状态来创建包括不同文字列的答复。例如,即使在用户讲出的话语的语

言分析结果是“asu wa hare”的情况下,如果话语的音高朝向词尾上升,则该话语也可变成询问(疑问句)“asu wa hare(desuka)?”(日语罗马字,意思是“明天是晴天吗?”)。因此,答复创建部110访问外部服务器以获取答复所需的天气信息,并且如果所获取的天气信息指示“hare”(日语罗马字,意思是“晴天”),则输出语音序列“hai”(日语罗马字,意思是“是”),或者如果所获取的天气信息指示不是“hare”,则输出语音序列“iie”(日语罗马字,意思是“不”)。此外,即使如上所述用户的话语的语言分析结果是“asu wa hare”,如果话语的音高变化是平坦的或者朝向词尾下降,则话语也可变为比如“asu wa hare(kâ)”的独白。因此,答复创建部110从答复数据库124读出或者输出比如“sou desune”的非正式反馈的语音序列。例如,如果音高变化量超过阈值,则答复创建部110确定话语的音高朝向词尾上升,而如果音高变化量等于或小于该阈值,则答复创建部110确定话语的音高是平坦的或者朝向词尾下降。

[0263] 在图28的步骤Sa16,语音控制部109B以与图4的步骤Sa16 相同的方式来识别从答复创建部110供应的语音序列的词尾的音高(初始音高)。

[0264] 随后,在步骤Sa17a,语音控制部109B基于音高数据和从非语言分析部107A供应的指示音高变化状态的数据按如下方式确定要对语音序列的音高做出的改变的内容。更具体地,如果用户的话语的音高朝向词尾上升,则语音控制部109B确定以如下方式改变整个语音序列的音高:使语音序列中所定义的词尾的初始音高呈现出从音高数据所指示的音高向下五度的预定的音程关系。另一方面,如果用户的话语的音高是平坦的或者朝向词尾下降,则语音控制部109B确定将整个语音序列的全部音高改变为从音高数据所指示的音高向下五度的音高。

[0265] 下面关于具体示例描述在第九实施例中执行的话语的音高、音高变化和语音序列的改变。图30的(b)的左边区域示出了用户给出的示例话语,其更具体地示出了其中用户的话语的语言分析结果是“asu wa hare”(日语罗马字,意思是“明天是晴天。”),并且其中如该部分中所示的按每个声音(音节)用音符指示音高的一个示例。注意,虽然如图29的(b)中所示的话语具有音高波形,但是该话语的音高用音符示出以便于描述。由于在示出示例中话语的音高朝向词尾上升,因此答复创建部110确定用户的话语是询问(疑问句)。因此,如果响应于该话语所获取的天气信息表明“hare”,则答复创建部110输出语音序列“hai”,或者如果所获取的天气信息表明不是“hare”,则答复创建部110输出语音序列“iie”。

[0266] 图30的(a)示出了答复“hai”(日语罗马字)的语音序列的基本语音的一个示例,其中按每个声音(音节)分配音符以定义该基本语音的每个词汇(音节或音素)的音高和产生时机。

[0267] 通过语音控制部109B按如下方式来改变这种基本答复语音序列。即,如果图30的(b)的左边区域指示的话语中由参考标记A 指示的词尾处的区间“re”的音高被音高数据指示为“G”,则语音控制部109B以如下方式改变整个语音序列的各音高:答复“hai”中由参考标记B指示的词尾处的区间“i”的音高变为从音高“G”向下五度。虽然关于答复“hai”已描述了语音序列改变,但是其他答复语音的整个语音序列的各音高以类似于上述方式的方式进行改变。

[0268] 在如上所述的用户的话语的语言分析结果是“asu wa hare”并且如图31的(b)的左边区域中所示话语的音高变化是平坦的情况下,答复创建部110确定话语是独白之类。因

此,答复创建部110输出例如之前所述的“sou desune”(日语罗马字,意思是“恩,……”)的语音序列。图31的(a)示出了“sou desune”的语音序列的基本语音的一个示例。这种基本语音序列由语音控制部109B按如下方式进行改变。

[0269] 即,如果图31的(b)的左边区域中指示的话语的中由参考标记A指示的词尾处的区间“re”的音高被音高数据指示为“G”,则语音控制部109B将答复语音“sou desune”(包括参考标记B指示的词尾处的“ne”)的全部音高改变为“C”,其是从“G”向下五度的音高(参见图31的(b)的右边区域)。

[0270] 同样在这种情况下,由于对答复的语音进行合成使得作为非正式反馈的答复的词尾的音高呈现出从词尾的音高向下五度的关系,因此可以给用户好似正在与人进行对话的良好印象,而不会给用户不自然的感觉。此外,根据本实施例,即使对于话语的相同的语言分析结果,也可以根据话语中朝向词尾的音高变化来创建不同答复。此外,如果话语的音高是平坦的,则针对该话语的非正式反馈的音高也是平坦的;即,原始语音序列中所定义的音高变化状态也被改变。因此,本实施例可以给予用户好似该用户正在与人而非与机器进行对话的印象。

[0271] <利用语音波形数据的修改例>

[0272] 类似上述第八实施例,第九实施例可以以与上述第二实施例相类似的方式来进行修改。即,图27的答复创建部110和语音合成部 112的组合(即,利用语音合成技术的语音产生部)可以被替换为如图8或图16或图19所示的答复语音输出部113和后处理部114的组合(即,利用波形数据的语音产生部),并且可以从答复语音输出部 113输出例如WAV格式的语音波形数据。

[0273] <答复的音高变化状态、答复的平均音高等>

[0274] 在上面描述了第九实施例被构造为对原始语音序列中所定义的答复的音高变化状态和平均音高进行改变,作为对整个语音序列的各音高进行移动使得答复的词尾等的音高变为例如从询问的词尾等的音高向下五度的结果,或者作为使语音序列的各音高平坦化的结果。但是,第九实施例不限于如此构造并且可以修改为对基本语音序列的音高变化状态进行改变。例如,对于其中话语的音高朝向词尾上升的音高变化状态,第九实施例可以改变基本语音序列的音高变化状态使得答复的音高朝向词尾下降,然而,对于其中话语的音高朝向词尾下降的音高变化状态,第九实施例可以改变基本语音序列的音高变化状态使得答复的音高朝向词尾上升。此外,第九实施例可以构造为对基本语音序列的全部或部分音高进行改变使得整个答复的平均音高根据话语的词尾等的音高或话语的音高变化而改变。

[0275] <话语的音量和音量变化、答复的音量和音量变化>

[0276] 以上描述了第九实施例被构造为利用话语的音量变化来识别话语的词尾。对于话语的这种音量,可以想到各种应用作为不同于音高信息的非语言信息。例如,第九实施例可以被构造为根据话语的平均音量来控制答复的合成语音的音量。此外,第九实施例可以被构造为根据话语的音量变化(音量包络)来控制答复的音量变化。

[0277] <对话的内容>

[0278] 以上描述了第九实施例被构造为在语音合成设备10以合成的语音输出针对用户的话语的答复时的时间点处终止其处理。但是,在人与人之间的实际对话中,话语和答复通常重复进行,而不是对话仅仅以话语和针对该话语的答复而结束,这种重复的数量根据话

语和答复的含义而增加或减少。因此,如图32所示,可以如下方式对第九实施例进行修改:语言分析部108不仅对用户的话语而且对答复创建部 110所创建的答复执行语言分析并将语言分析结果供应至语音控制部109,并且语音控制部109根据语言分析结果对答复的词尾等的音高、答复的音高变化状态、答复的平均音高等进行控制。

[0279] 注意,关于第一实施例和第二实施例的各种应用示例和/或修改例也可应用于本发明的第九实施例。

[0280] 总之,上述第九实施例的特征在于包括:非语言分析部(107A),其对话语中的音高变化进行分析;和控制部(语音控制部109B),其根据所分析的话语中的音高变化来执行用于对语音产生部(即,语音合成部112,或答复语音产生部(113)和后处理部(114)的组合)产生的答复的语音的音高进行改变的控制。

[0281] <第十实施例>

[0282] 下面参照图33描述本发明的语音合成设备的第十实施例,其特征在于,根据话语的说话者或讲话者的属性或答复语音的属性来修改用于确定答复语音中的第二区间的音高的规则。图33是本发明的语音合成设备10的第十实施例的功能框图。用与图1中相同的参考标记示出与图1中的块实现基本相同功能的图33中的块,并且将不对其进行描述以避免不必要的重复。图33的实施例除了包括图1示出的构造之外还包括语音控制部109C。

[0283] 类似于上述各实施例中的对应部分,第十实施例中的语音合成部112通过根据给定规则(即,以如下方式:使答复的第二区间的音高与询问或话语的词尾的音高呈现出预定的音程关系)对答复的第二区间的音高进行控制来对答复的语音进行合成。此外,第十实施例中的语音合成部112被构造为产生具有基于给定中介属性的特点的答复的语音。中介属性指示其语音将要被语音合成设备10合成的虚拟人的信息。即,中介属性是指示人的数据,用于定义假设哪种人来合成答复的语音。为了简化描述,这里假设中介属性是定义性别的数据。本实施例以如下方式构造:由用户经由属于语音合成设备10的操作输入部来选择或设置期望的中介属性,并且指示所选择或设置的中介属性的信息被供应至语音合成部112并合成具有基于该中介属性的特点的答复的语音。

[0284] 为了对答复语音序列的音高进行控制而应用于语音合成部112 中的上述规则通过语音控制部109C来确定。作为规则的一个示例默认选项(默认规则),规定了:语音控制部109C应当将答复的词尾的音高进行移动从而与话语(询问)的词尾的音高呈现出预定的音程关系,更具体地,从话语(询问)的词尾的音高向下五度。但是,在一种替选方式中,语音控制部109C可以将答复的词尾的音高移动至具有不同于上述的从话语(询问)的词尾的音高向下五度的音程关系的音程关系。然而,如果坚持使用默认规则,则通过语音合成而创建的答复可能不期望地变得不自然。因此,语音控制部109C被构造为必要时根据话语讲话者的属性(话语讲话者属性)和中介属性对默认规则进行修改,从而确定要应用于语音合成部112的规则。话语讲话者属性是输入话语(询问)的讲话者的属性,在本实施例中,其是讲话者(用户)的性别。可以使用例如在充当语音合成设备10的终端设备中登记的用户的个人信息来作为对话语讲话者属性进行定义的数据。

[0285] 对于语音合成,语音合成部112使用在语音库128中登记的语音片段数据。将被用作语音素材的定义了各种语音片段(诸如各单独音素和音素至音素的过渡部分)的波形的语音片段数据预先编译作为针对多个中介属性中的每一个的数据库。更具体地,语音合成

部112 利用由选定或设定的中介属性所定义的语音片段数据来产生语音信号,更具体地,通过将语音序列的各单独声音(音节)的语音片段数据进行组合并将所组合的语音片段数据修改为连续地相互连接来产生语音信号。

[0286] 图34是语音合成设备10的第十实施例中执行的处理的流程图。图34被示出为图4的修改例,并且与图4中的步骤执行实质相同的操作的步骤用相同参考标记示出并将不在此处进行描述以避免不必要的重复。

[0287] 在图34的步骤Sa15a和Sa15b处执行与图4的步骤Sa15和图 21的步骤Sa15a和Sa15b相类似的操作。此外,在步骤Sa16,以与图14的步骤Sa16处相同的方式,执行对答复语音序列中的词尾的音高(初始音高)进行识别的操作。此外,以与图14的步骤Sa13处相同的方式,对检测到的说话区间中的话语的语音信号执行分析以识别话语中第一区间(词尾)的音高。指示由此识别出的音高的音高数据被供应至语音控制部109C。

[0288] 在步骤Sa27,由语音控制部109C基于中介属性和话语讲话者属性来执行规则确定处理,从而确定要应用的规则。图35是示出规则确定处理的细节的流程图。

[0289] 首先,在图35的步骤Sb11,语音控制部109C获取指示话语讲话者属性的数据和指示中介属性的数据。随后,在步骤Sb 12,语音控制部109C基于所获取的数据来确定话语讲话者属性(即,用户属性)是否为女性。如果话语讲话者属性是女性(步骤Sb12处判定为是),则语音控制部109C对默认规则进行修改使得答复的词尾的音高被移动至从音高数据所指示的音高向下六度(而非默认的五度)的音高,即,被移动至与音高数据所指示的音高处于协和音程关系且与音高数据所指示的音高的默认关系向下一级的音高。以这种方式,从由默认规则预设的音高下降的音高在步骤Sb13被确定为答复的词尾的音高。术语“级(rank)”不具有任何音乐含义并且仅为了便于描述而在这里使用。即,在本实施例中,在将从音高数据所指示的音高向下五度的音高设置为参考音高的情况下,从参考音高降一级的音高指的是从音高数据所指示的音高向下六度(大六度)的音高,从参考音高又降一级的音高指的是从音高数据所指示的音高向下八度的音高。此外,从参考音高升一级的音高指的是从音高数据所指示的音高向下(低于)三度(大三度)的音高,而从参考音高又升一级的音高指的是从音高数据所指示的音高向上四度的音高。

[0290] 另一方面,如果话语讲话者属性不是女性(步骤Sb12处判定为否),则语音控制部109C在步骤Sb14进一步确定话语讲话者属性是否为男性。如果话语讲话者属性是男性(步骤Sb14处判定为是),则语音控制部109C对默认规则进行修改以使得答复的词尾的音高被移动至从音高数据所指示的音高向下三度的音高。以这种方式,在步骤Sb15将从默认规则所预设的音高上升的音高确定为答复的词尾的音高。如果话语讲话者属性是中性或者话语讲话者属性还未被登记(步骤Sb14处判定为否),则语音控制部109C跳过步骤Sb13或Sb15的操作并且使用未修改形式的默认规则。

[0291] 随后,在步骤Sb16,语音控制部109C确定中介属性是否为女性。如果中介属性是女性(步骤Sb16处判定为是),则语音控制部109C 在步骤Sb17对修改后的默认规则(或者未修改的默认规则)进行修改以使得答复的词尾的音高被上移一级。例如,如果在上述步骤Sb13处已修改了默认规则使得答复的词尾的音高下移一级或从音高数据所指示的音高向下六度,则语音控制部109C在步骤Sb17将修改后的规则返回至原始默认规则以使得答复的词尾的音高从音高数据所指示的音高向下移动五度。此外,如果在上述步骤Sb15处已修改了默

认规则使得答复的词尾的音高上移一级或从音高数据所指示的音高向下三度,则语音控制部109C在步骤Sb17进一步对修改后的默认规则进行修改以使得答复的词尾的音高又上移一级或从音高数据所指示的音高向上四度。注意,如果如上所述跳过了步骤Sb13或Sb15 的操作,则语音控制部109C在步骤Sb17对默认规则进行修改以使得答复的词尾的音高上移一级或从音高数据所指示的音高向下三度。

[0292] 如果中介属性不是女性(步骤Sb16处判定为否),则语音控制部109C在步骤Sb18进一步确定中介属性是否为男性。如果中介属性是男性(步骤Sb18处判定为是),则语音控制部109C在步骤Sb19 进一步对修改后的默认规则进行修改以使得答复的词尾的音高下移至向下一级的音高。例如,如果默认规则已被修改使得答复的词尾的音高下移一级或从音高数据所指示的音高向下六度,则语音控制部 109C在步骤Sb19进一步修改默认规则以使得答复的词尾的音高被移动至从参考音高再向下一级的音高或从音高数据所指示的音高向下八度的音高。此外,如果已修改了默认规则使得答复的词尾的音高被移动至向上一级的音高或从音高数据所指示的音高向下三度的音高,则语音控制部109C将修改后的默认规则返回至原始默认规则以使得答复的词尾的音高被移动至最初规定的向下五度的音高。此外,如果跳过了步骤Sb13或Sb15的操作,则语音控制部109C在步骤Sb19 修改默认规则,使得答复的词尾的音高被移动至从参考音高向下一级的音高或者从音高数据所指示的音高向下六度的音高。

[0293] 如果中介属性是中性或者如果中介属性还未被登记(步骤Sb18 处判定为否),则语音控制部109C跳过步骤Sb17或Sb19的操作。在完成步骤Sb17或Sb19的操作之后,或者在跳过步骤Sb17或Sb19 之后,处理返回至图34的步骤Sa28。通过酌情修改的默认规则来确定要应用的规则。注意,可以基于中介属性或话语讲话者属性中的至少一个来执行默认规则的修改(即,规则的确定)。

[0294] 返回来参考图34,语音控制部109C在步骤Sa28确定通过应用在步骤Sa27处确定的规则(或默认规则)来改变从答复创建部110 供应的语音序列。更具体地,如果所确定的规则规定了答复中的词尾的音高应当被移动至例如从音高数据所指示的音高向下三度的音高,则语音控制部109C以如下方式对语音序列的全部音高进行移动:使从答复创建部110供应的语音序列中所定义的答复中的词尾的音高呈现出从音高数据所指示的音高向下三度的音程关系。在步骤Sa18,语音控制部109C根据所确定的内容控制语音合成部112进行的语音合成。由此,语音合成部112根据语音控制部109C所确定的规则来控制语音序列的各音高并从而合成和输出具有受控音高的答复语音信号。

[0295] 下面关于一些具体示例来描述话语的音高、语音序列的基本音高以及改变后的语音序列的音高。图36的(b)的左边区域示出了用户讲出的话语的一个示例。在图36的示出示例中,如左边区域中所示,该话语的语言分析结果为“asu wa hare desuka?”(日语罗马字,意思是“明天是晴天吗?”)并且其中利用音符将音高分配给话语的各独立声音。虽然该话语实际上具有类似于图5的(b)所示音高波形的音高波形,但是为了便于说明用音符来示出该话语的各音高。在这种情况下,与之前所述方式相同,如果响应于该话语而获取的天气信息指示“hare”,则答复创建部110输出语音序列“hai”,或者如果所获取的天气信息指示不是“hare”,则答复创建部110 输出语音序列“iie”。

[0296] 图36的(a)示出了语音序列“hai”(日语罗马字)的基本音高的一个示例,而图36的

(b) 的右边区域示出了根据默认规则控制了其音高的答复语音序列的一个示例。即,如果将要应用默认规则,则答复创建部110输出的语音序列按如下方式被语音控制部109C改变。即,如果图36的 (b) 的左边区域中指示的话语中由参考标记A 所指示的词尾处的区间“ka”的音高由音高数据指示为“E”,则语音控制部109C改变整个语音序列的音高使得答复“hai”中由参考标记B所指示的词尾处的区间“i”的音高变为音高“A”,其为从音高“E”向下五度的音高(参见图36的 (b) 的右边区域)。

[0297] 注意,在本实施例中,在三种情况下应用默认规则:当在步骤 Sb12、Sb14、Sb16和 Sb18的每一个中均做出“否”判定时;当在步骤Sb12做出“是”判定并且在步骤Sb16做出“是”判定时;以及当在步骤Sb12做出“否”判定,在步骤Sb14做出“是”判定,并且在步骤Sb18做出“是”判定时。

[0298] 在讲出的话语如图36的 (b) 的左边区域所示的情况下并且如果指示移动至例如从音高数据所指示的音高向下六度的音高的修改后的规则将被应用,则答复创建部110输出的语音处理按如下方式被语音控制部109C改变。即,语音控制部109C改变整个语音序列的各音高,使得答复“hai”中由参考标记B所指示的词尾处的区间“i”的音高被改变为从音高“E”向下六度的音高“G”(参见图37的右边区域)。

[0299] 注意,在本实施例中,在两种情况下应用指示“向下六度”的规则:当在步骤Sb12做出“是”判定,并且在步骤Sb16和步骤Sb18 做出“否”判定时;以及当在步骤Sb12和Sb14做出“否”判定,在步骤Sb16做出“否”判定,并且在步骤Sb18做出“是”判定时。

[0300] 在讲出的话语如图36的 (b) 的左边区域所示的情况下,并且如果将要应用指示移动至例如从音高数据所指示的音高向下八度的音高的修改后的规则,则由答复创建部110输出的语音序列按如下方式被语音控制部109C改变。即,语音控制部109C改变整个语音序列的各音高使得答复“hai”中由参考标记B所指示的词尾处的区间“i”的音高被改变为从音高“E”向下八度(一个八度)的音高“E”(参见图38的右边区域)。注意,在本实施例中,只在一种情况下应用指示“向下八度”的规则,即,当在步骤Sb12做出“是”判定,在步骤Sb16做出“否”判定并且在步骤Sb18做出“是”判定时。

[0301] 在讲出的话语如图36的 (b) 的左边区域所示的情况下并且如果将要应用指示移动至从音高数据所指示的音高向下三度的音高的修改后的规则,则答复创建部110输出的语音序列按如下方式被语音控制部109C改变。即,语音控制部109C改变整个语音序列的各音高,使得答复“hai”中由参考标记B所指示的词尾处的区间“i”的音高被改变为从音高“E”向下三度的音高“C”(参见图39的右边区域)。注意,在本实施例中,在两种情况下应用指示“向下三度”的规则:当在步骤Sb12做出“否”判定,在步骤Sb14做出“是”判定并且在步骤Sb16和 Sb18做出“否”判定时;以及当在步骤Sb12和Sb14 做出“否”判定并且在步骤Sb16做出“是”判定时。

[0302] 在讲出的话语如图36的 (b) 的左边区域所示的情况下并且如果将要应用指示移动至从音高数据所指示的音高向上四度的音高的修改后的规则,则答复创建部110输出的语音序列按如下方式被语音控制部109C改变。即,语音控制部109C改变整个语音序列的各音高,使得答复“hai”中由参考标记B所指示的词尾处的区间“i”的音高被改变为从音高“E”向上四度的音高“A”(参见图40的右边区域)。注意,在本实施例中,只在一种情况下应用指示“向上四度”的规则,即,当在步骤Sb12做出“否”判定,在步骤Sb14做出“是”判定并且在步骤

Sb16做出“是”判定时。

[0303] 以上关于答复包括词汇“hai”的情况对本实施例进行了描述。但是,对于包括不同于“hai”的其他词汇的答复也一样,以与前述方式类似的方式,根据基于中介属性和话语讲话者属性中的至少一个所确定的规则来改变整个答复语音序列的各音高。

[0304] 在应用规定了答复的词尾的音高应当是从话语的词尾的音高向下五度的默认规则的情况下,在本实施例中以如下方式合成答复的语音:如果讲话者属性是女性,则答复的词尾的音高下降一级,而如果讲话者属性是男性,则上升一级。此外,在应用规定了答复的词尾的音高应当是从话语的词尾的音高向下五度的默认规则的情况下,在本实施例中对答复的语音进行合成以使得如果中介属性是女性,则将答复的词尾的音高上升一级,而如果中介属性是男性,则下降一级。如上所述,由于答复的各音高根据讲话者属性和中介属性而改变,因此本实施例可给用户某种程度的新鲜感和乐趣。

[0305] <利用语音波形数据的修改例>

[0306] 如上述第八实施例和第九实施例一样,第十实施例可以以与上述第二实施例类似的方式进行修改。即,答复创建部110和语音合成部112的组合(即,利用语音合成技术的语音产生部)可以被替换为如图8或图16或图19所示的答复语音输出部113和后处理部114 的组合(即,利用波形数据的语音产生部),并且可以从答复语音输出部113输出例如WAV格式的语音波形数据。

[0307] <讲话者属性>

[0308] 以上将第十实施例描述为利用在充当语音合成设备10的终端设备中登记的用户个人信息作为讲话者属性。但是,作为一种替代方式,可以在语音合成设备10中检测讲话者属性。例如,可以对用户的话语执行音量分析、频率分析等,并且随后将这种被分析的用户的话语和与预先存储的各种性别和年龄的组合相对应的模式进行比较,从而检测出具有高相似度的一种模式的属性作为讲话者属性。如果无法检测到这种讲话者属性,则在图35的步骤Sb12和步骤Sb14做出“否”判定。

[0309] <中介属性>

[0310] 虽然以上关于中介属性为性别的情况对第十实施例进行了描述,但是中介属性可以包括三种或三种以上属性(诸如性别、年龄等) 的组合。

[0311] <非正式反馈的重复、非正式反馈的输出时机等>

[0312] 当从讲话者的性别的角度来看人与人之间的对话时,可能会根据讲话者属性观察到以下特定趋势。为了便于描述,假设女性通常倾向于重视氛围和调和并且在对话中观察到使氛围活跃的趋势。更具体地,假设观察到频繁使用非正式反馈、重复非正式反馈、缩短从话语到答复的时长等趋势。进一步假设女性肯定也对以合成的语音输出针对话语的答复的语音合成设备10抱以这种期望。因此,如果讲话者属性是女性,则语音控制部109C可以通知答复创建部110该结果以使得答复创建部110增加将答复创建为针对话语的非正式反馈的频率或者重复地输出同一非正式反馈的语音序列。此外,语音控制部 109C可以控制语音合成部112以相对加快从用户的话语的结尾到开始输出针对该话语的答复的时间点的时间。

[0313] 另一方面,男性通常倾向于重视对话中的内容、逻辑性、个性等。更具体地,为了便于描述,假设男性倾向于不会不必要地给出非正式反馈、根据情形敢于不给出答复(敢于沉

默)、延长从话语到答复的时长等。因此,如果讲话者属性是男性,则语音控制部109C可以通知答复创建部110该结果以使得答复创建部110降低将答复创建为针对话语的非正式反馈的频率或者有时在一定几率下不给出答复。此外,语音控制部109C可以控制语音合成部112以相对减慢从用户的话语的结尾到开始输出针对该话语的答复的时间点的时间。

[0314] 此外,作为在答复的词尾的音高从根据默认规则而预先确定的音高降低时要应用的一个条件,可以在图35的步骤Sb13将话语的词尾的音高应当等于或大于第一阈音高(频率)的条件(参见步骤Sb13的块中的标记※)添加到讲话者属性是女性的条件。这是为了避免答复(包括合成的语音)在女性讲话音高高的情况下变得不自然地高。类似地,作为在答复的词尾的音高从根据默认规则而预先确定的音高升高时要应用的一个条件,可以在图35的步骤Sb15处将话语的词尾的音高应当等于或小于第二阈音高的条件(参见步骤Sb15的块中的标记※)添加到讲话者属性是男性的条件。这是为了避免答复(包括合成的语音)在男性讲话音高低的情况下变得不自然地低。

[0315] 注意,关于第一实施例和第二实施例的各种应用示例和/或修改例也可应用于本发明的第十实施例。

[0316] 简而言之,上述第十实施例的特征在于:语音产生部(即,语音合成部112,或答复语音输出部113和后处理部114的组合)被构造为根据给定规则将第二区间的音高与所分析的第一区间的音高相关联,并且利用基于给定中介属性的特点来产生答复的语音。第十实施例还包括控制部(语音控制部109C),其基于中介属性和话语的讲话者的属性中的至少一个确定上述规则。

[0317] <第十一实施例>

[0318] 下面参照图41描述本发明的语音合成设备的第十一实施例,其特征在于:在答复语音的可听地产生或发声之后,响应于经由语音输入部接收到另一话语的语音,对用于将答复语音的第二区间的音高与所分析的第一区间的音高相关联的规则进行更新。即,第十一实施例的特征在于:鉴于响应于机器讲出的答复而由用户讲出的另一话语,酌情改变要应用于下一答复语音的产生的规则。图41是本发明的语音合成设备10的第十一实施例的功能性框图。用与图1中相同的参考标记示出与图1中的块实现基本相同功能的图41中的块,并且将不对其进行描述以避免不必要的重复。图41的实施例除了包括图1示出的构造之外还包括语音控制部109D和管理数据库127。

[0319] 类似于上述实施例中的每一个中的对应部分,第十一实施例中的语音合成部112通过根据预定规则(即,使得答复的第二区间的音高与询问或话语的第一区间的音高呈现出预定的音程关系)对由答复创建部110所创建的答复语音序列的各音高进行控制来对答复的语音进行合成。下面将第十一实施例中采用的给定规则称为“音高规则”。在答复语音的可听地产生或发声之后响应于经由语音输入部102接收到的另一话语的语音,语音控制部109D通过对应用至语音合成部112的规则(音高规则)进行更新来控制语音合成部112进行的语音合成。

[0320] 如上所述,答复的第二区间的音高相对于询问或话语的第一区间的音高应当呈现出何种关系以让用户感到舒服并使得对话逼真在用户与用户之间存在不同。因此,在第十一实施例中,将评估时段设置为操作时段,并且利用多个音高规则来合成针对同一话语的各答复的语音。随后,在评估时段结束时,其中实现最逼真对话的一个音高规则(即,利用该

音高规则使得对话最逼真)被设置为要应用的音高规则,使得可以在随后的语音合成中反映由此设置的音高规则。

[0321] 由语音控制部109D管理的管理数据库127除了别的之外还存储其中音高规则和指示对话的逼真度的指标彼此一一关联的表格(指标表)。图42是示出指标表中的存储内容的一个示例。如图42所示,针对每个音高规则将话语的数量和应用的数量彼此关联。这里,各音高规则的每个都规定了答复的词尾的音高相对于话语的词尾的音高应当呈现出何种关系,比如向上四度、向下三度、向下五度、向下六度和向下八度,如图中所示。

[0322] 此外,“话语的数量”意思是在语音合成设备10合成了针对用户讲出的话语的答复的语音并且随后用户在预定时间段内讲出另一话语的情况下用户讲出的话语的计数数量。话句话说,在评估时段期间,即使在答复的语音已经被语音合成设备10合成时,也可以想到的是用户没有讲出另一话语或者在预定时间段过去之后讲出另一话语(如有);这样的另一话语不被作为话语的数量进行计数。“应用的数量”指的是在评估时段期间相应音高规则被应用的次数。因此,通过比较将话语数量除以相应应用数量而计算出的值,用户可以了解哪个音高规则实现了最大数量的针对答复的话语,即,哪个音高规则实现了最逼真的对话。注意,即使已通过应用音高规则中的任一个合成了答复的语音时,用户可能有时也在预定时间段内不针对该答复讲出话语。这是因为应用的数量超过了话语的数量,如图42的示出示例中那样。

[0323] 图43是示出由CPU执行的应用程序所设置的操作时段的一个示例的图。在本实施例中,如图43所示,在操作时段中,规则固定时段和上述评估时段交替重复进行。这里,规则固定时段是根据在评估时段结束时设置的一个音高规则来合成答复的语音的时段。作为一个示例,在其中一个规则固定时段设置的这种音高规则是如图中白色三角所指示的“向下五度”。

[0324] 另一方面,评估时段是用于基于通过将音高规则应用于用户讲出的话语而进行的答复的语音合成来评价出多个音高规则中实现了最逼真的对话的音高规则。虽然本实施例被构造为使得规则固定时段和上述评估时段在操作时段中如图43所示那样交替重复,但是其也可以被构造为例如只响应于用户的指令而转移至评估时段。

[0325] 图44是在语音合成设备10的第十一实施例中执行的语音合成处理的流程图。与规则固定时段和评估时段无关地执行该语音合成处理。图44被示出为图4的修改例,并且与图4中的步骤执行实质相同的操作的步骤用与图4中相同的参考标记示出并将不进行描述以避免不必要的重复。

[0326] 在图44的步骤Sa15a和Sa15b处执行与图4的步骤Sa15或图21的步骤Sa15a和Sa15b类似的操作。在图44的步骤Sa16,以与图4的步骤Sa16相同的方式执行用于识别从答复创建部110供应的语音序列中的第二区间(词尾)的音高(初始音高)的操作。此外,在图44的步骤Sa13a,以与图4的步骤Sa13相同的方式分析话语中的检测到的说话区间的语音信号以识别话语中的第一区间(词尾)的音高。随后,指示识别出的音高的数据被供应至语音控制部109D。

[0327] 在步骤Sa29,语音控制部109D确定当前时间点是否处于规则固定时段之一内。如果当前时间点处于规则固定时段之一内(在步骤Sa29判定为是),则语音控制部109D在步骤Sa30应用在规则固定时段之前的评估时段中设置的音高规则。如果当前时间点不处于规

则固定时段之一内而是处于评估时段之一内(步骤Sa29判定为否),则语音控制部109D在步骤Sa31选择例如以下总计三个音高规则中的任一个:在当前评估时段之前的评估时段中设置的音高规则以及在指标表中上下紧挨着(即,在上下方向上夹住)该音高规则的两个规则,然后应用由此选择的音高规则。更具体地,如果所设置的音高规则是图42中的白色三角标记所示出的“向下五度”,则语音控制部109D随机或按预定顺序选择以下三个规则中的任一个:“向下五度”的音高规则以及在指标表的竖直或上下方向上夹住“向下五度”的“向下三度”和“向下六度”的音高规则。

[0328] 在下一个步骤Sa32,语音控制部109D指示语音合成部112对答复进行音高改变以使得答复的第二区间(词尾)的音高与在步骤Sa13 识别出的话语的第一区间(词尾)的音高呈现出由所应用的音高规则规定的预定的音程关系。根据这样的指示,语音合成部112以如下方式改变整个语音序列的各音高:使答复的词尾的音高移动至由所应用的音高规则规定的音高。在本实施例中,在通过语音合成(即,以合成的语音)输出一个答复之后处理返回至步骤Sa11,从而允许用户在该答复之后讲出另一话语(即,从而接收另一话语)。如果用户不想继续与机器进行对话,则可响应于用户的明确操作(例如,用户对软件按钮的操作)使当前语音合成处理结束。

[0329] 图45是示出表更新处理的流程图。该表更新处理(其独立于图 44的语音合成处理而执行)主要被设计来在评估时段对指标表(参见图42)进行更新,从而设置要在规则固定时段应用的音高规则。

[0330] 在步骤Sb21,语音控制部109D确定当前时间点(当前时间)是否处于评估时段之一内。如果当前时间点未处于评估时段之一内(步骤Sb21处判定为否),则语音控制部109D将表更新处理返回至步骤 Sb21。另一方面,如果当前时间点处于评估时段之一内(步骤Sb21处判定为是),则语音控制部109D在步骤Sb22进一步确定是否已输出通过语音合成部112进行的语音合成所创建的任何答复。如果答复还未输出(步骤Sb22处判定为否),则语音控制部109D将表更新处理返回至步骤Sb21。由此,不执行表更新处理中的随后操作,除非当前时间点处于评估时段之一内并且已输出任一答复。另一方面,如果答复已输出(步骤Sb22处判定为是),则语音控制部109D在步骤Sb23 进一步确定在该答复输出之后的预定时间段(例如,5秒)内是否存在任何用户话语。可以通过语音控制部109D对在答复输出之后的预定时间段内是否存在被供应的音高数据进行检查来确定在预定时间段内是否存在任何用户话语。

[0331] 如果在答复的输出之后的预定时间段内存在任何用户话语(步骤 Sb23处判定为“是”),则语音控制部109D在步骤Sb24识别已被应用至答复的语音合成的音高规则,以对指标表进行更新。通过如下方式容许进行对这种音高规则的识别:在上述步骤Sa31处选择音高规则时将选定的音高规则和选定的时间信息彼此关联地存储到管理数据库 127中,并随后搜索与最新或最近的时间信息相关联的音高规则。在步骤Sb25,语音控制部109D在指标表中将应用至答复的语音合成的音高规则的各项(话语的数量和应用的数量)增加1。

[0332] 另一方面,如果在答复的输出之后的预定时间段内不存在用户话语或者在预定时间段过去之后才讲出话语(步骤Sb23处判定为“否”),则语音控制部109D在步骤Sb26以与步骤Sb24相同的方式识别已被应用至答复的语音合成的音高规则。但是,在这种情况下,语音控制部 109D在步骤Sb27只将指标表中的应用至答复的语音合成的音高规则的应用的数量

增加1,这是因为认为不存在用户话语。

[0333] 接着,在步骤Sb28,语音控制部109D确定当前时间点是否是评估时段的结束时间。如果当前时间点不是评估时段的结束时间(步骤 Sb28处判定为“否”),则语音控制部109D将处理返回到上述步骤 Sb21以为在答复的输出之后用户讲出话语时的下一次做准备。另一方面,如果当前时间点是评估时段的结束时间(步骤Sb28处判定为“是”),则语音控制部109D在所讨论的评估时段中针对三个音高规则将通过将话语的数量除以相应的应用的数量而计算出的数值进行比较,从而语音控制部109D将各音高规则中实现最逼真对话的一个音高规则设置为将要在该评估时段之后的规则固定时段中应用的音高规则(步骤 Sb29)。例如,如果在步骤Sb28的操作中,评估时段中的三个音高规则分别是向下三度、向下五度、向下六度,并且话语的数量和应用的数量是如图42所示的值,则在规则固定时段中要应用的音高规则从预先设置的“向下五度”音高规则改变为由黑色三角标记所示出的“向下三度”音高规则。然后,语音控制部109D将在该评估时段中评估的三个音高规则中每一个的话语的数量和应用的数量清除,并随后将处理返回至步骤Sb21以便在下一个评估时段中执行类似操作。

[0334] 如上所述,本实施例被构造为:通过在评估时段中应用不同音高规则中的任一个来创建答复;如果在预定时段内存在针对答复的任何用户话语,则对话语的数量和所应用的音高规则的应用的数量进行更新;以及如果针对答复的用户话语不在预定时段内,则只对所应用的音高规则的应用的数量进行更新。随后,在评估时段结束时间处,在下一个规则固定时段中设置和应用实现了最逼真的对话的音高规则。

[0335] 第十一实施例中的话语的音高、语音序列的基本音高和语音序列的改变后的音高的具体示例可以与图36至图40中的那些类似。即,如果将“向下五度”应用为用于确定针对如图36的(b)的左边区域中所示的话语“asu wa hare desuka?”的答复“hai”的各音高的规则,则设置如图36的(b)的右边区域所示的音高模式。如果“向下六度”被应用为所述规则,则设置如图37的右边区域所示的音高模式。此外,如果“向下八度”被应用为所述规则,则设置如图38的右边区域所示的音高模式。此外,如果“向下三度”被应用为所述规则,则设置如图39的右边区域所示的音高模式。此外,如果“向上四度”被应用为规则,则设置如图40的右边区域所示的音高模式。

[0336] 在第十一实施例中,要在规则固定时段中应用的音高规则是在该规则固定时段之前的评估时段中实现了最逼真的对话的音高规则。因此,在该规则固定时段中,可以容易地使对话变得逼真;简言之,用户容易讲出话语。此外,由于在各评估时段中设置了这样的音高规则,因此可以实现使用户舒服、可以让用户放松并且可以使对话逼真(即,可以将对话指引到更有生气的方向上)的情形。

[0337] <第十二实施例>

[0338] 已描述了第十一实施例被构造为:在评估时段应用多个音高规则、设置各音高规则中实现了最逼真的对话的任一音高规则并且在规则固定时段中使用由此设置的音高规则。然而,除了音高之外,能够使得对话逼真的因素的示例还包括从话语到答复的“停顿”,即,时间间隔。因此,下面描述本发明的第十二实施例,其不仅如第十一实施例那样基于对音高规则的设置来执行答复语音音高控制,而且还在评估时段输出具有不同停顿的答复、设置各停顿中实现了最逼真的对话的一个停顿并且应用由此设置的停顿以控制答复的停

顿。

[0339] 通过执行上述应用程序而在第十二实施例中构建的功能块大致与图41的第十一实施例中的那些类似。然而,在第十二实施例中,除了如图42所示的用于对音高规则进行评估的表以外,还使用如图 46所示的用于评估答复输出规则的表作为指标表。

[0340] 在用于评估答复输出规则的指标表(如图46所示)中,对于每个输出规则而言话语的数量和应用的数量彼此关联。这里,输出规则的每一个规定了例如从话语的结束(词尾)到答复的开始(词首)的停顿或时间间隔,并且这些输出规则以阶梯式方式指定了0.5秒、1秒、1.5秒、2.0秒和2.5秒。注意,这里的与各单独输出规则相关联的话语的数量和应用的数量与第十一实施例中的类似。

[0341] 第十二实施例中的处理可以类似于图44和图45中示出的处理,不同之处在于图44和图45中的“音高规则”应当被读作“音高规则和输出规则”等。更具体地,如果在图44中的步骤Sa30处当前时间点处于规则固定时段内,则语音控制部109D决定通过应用已经在该规则固定时段之前的评估时段中设置的音高规则和输出规则来合成语音。另一方面,如果在步骤Sa31当前时间点处于评估时段内,则语音控制部109D选择三个音高规则中的任一个,选择如下总共三个输出规则中的任一个:已经在规则固定时段之前的评估时段中设置的输出规则;以及在指标表(参见图46)中上下紧挨着(即,在竖直或上下方向上夹住)该设置的输出规则的两个输出规则,并应用由此选择的音高规则和输出规则。在步骤Sa32,被提供了音高数据的语音控制部109D指示语音合成部112合成针对话语的答复的语音以使得答复的词尾的音高与由所提供的音高数据所指示的音高呈现出如所应用的音高规则所确定的关系,并且使得从话语的词尾到开始输出答复时的时间点的时间间隔与由所应用的输出规则所确定的时间间隔一致。

[0342] 此外,为了更新所述两个指标表,语音控制部109D在图45的步骤Sb24和Sb26识别应用至答复的语音合成的音高规则和输出规则,并且在步骤Sb25将所应用的音高规则的两项和所应用的输出规则的两项中的每一项增加1。在步骤Sb27,语音控制部109D只将所应用的音高规则的应用的数量增加1并且只将所应用的输出规则的应用的数量增加1。如果当前时间点是评估时段的结束时间,则语音控制部109D在步骤Sb29在评估时段中设置各音高规则和各输出规则中实现了最逼真的对话的一个音高规则和一个输出规则。然后,在步骤Sb30,语音控制部109D将在评估时段中评估的各音高规则和输出规则的各项清除。

[0343] 通过第十二实施例,在评估时段中各音高规则和各输出规则中的实现了最逼真的对话的一个音高规则和一个输出规则被应用到该评估时段之后的规则固定时段中,可以利用有助于用户讲话的停顿来返回让用户舒服的良好印象的答复。例如,在语音合成设备10响应于如图17的(a)所示的用户的话语“asu no tenki wa?”(日语罗马字,意思是“明天天气如何?”)而输出答复“hare desu”(日语罗马字,意思是“明天是晴天”)的情况下,将从用户的话语的词尾“wa”(日语罗马字)到词首“ha”(日语罗马字)的时间间隔 Ta设置为有助于用户W使对话有生气的时间间隔。在这种情况下,尽管未特别示出,答复的词尾“su”(日语罗马字)的音高被设置为相对于用户的话语的词尾“wa”(日语罗马字)处于由可以有助于使对话有生气的音高规则所指定的关系。

[0344] 如上所述,在第十二实施例中,如第十一实施例那样,以如下方式合成答复的语音:使答复的词尾的音高与话语的词尾的音高呈现出协和音程关系。此外,第十二实施例利

用停顿来合成答复的语音,该停顿允许用户相比于第十一实施例更容易讲话,由此,第十二实施例可以使与用户的对话相比于第十一实施例更加逼真。

[0345] 虽然如上描述了第十二实施例被构造除了以与第十一实施例相同的方式执行答复的音高控制之外,还控制从话语到答复的“停顿”,但是其可以被构造为只控制停顿而不执行答复的音高控制。这种修改形式(执行停顿控制而不执行音高控制)的内容通常类似于图 45 的处理,不同之处在于图 45 中的“音高规则”应当读作“输出规则”,并且从以上关于第十二实施例的描述中本领域技术人员可以充分理解该修改形式的内容。

[0346] <第十三实施例>

[0347] 下面描述本发明的第十三实施例。首先,简要叙述第十三实施例的前提。使针对话语的答复的词尾的音高相对于话语的词尾的音高让人感到舒服的音高关系因人而异。特别是,由于女性和男性的话语的音高彼此差异很大(即,女性的音高通常高于男性的音高),因此女性和男性的话语可以给出极大不同的印象。此外,近年来,有时可以通过语音合成以预定性别和年龄的虚拟人物的语音来输出答复。据认为,如果针对话语进行答复的人物的语音改变并且尤其是如果人物的性别改变,则用户将获得与改变前接收到的印象不同的印象。因此,本发明的第十三实施例作为不同场景假设了用户的性别(女性和男性)和要合成的语音的性别的各种组合并提供针对各单独场景的指标表以使得可以使用各指标表中与用户的话语相对应的任一个指标表。

[0348] 图 47 示出了在第十三实施例中提供与用户的性别和要合成的语音的性别的各种组合相对应的指标表的示例。更具体地,在管理数据库 127 中准备了与用户的两种性别(即,女性和男性)以及要通过设备合成的答复语音的两种性别(即,女性和男性)对应的总计四个指标表。语音控制部 109D 以如下方式来选择这四个指标表中的任一个。

[0349] 即,语音控制部 109D 根据例如已登录到充当语音合成设备 10 的终端设备中的用户的个人信息来识别用户的性别。替代性地,语音控制部 109D 可以对用户的话语执行音量分析和频率分析,将所分析的用户的话语的音量和频率与预先存储的女性和男性的模式进行比较,并识别预先存储的模式中与用户的话语的音量和频率具有高相似度的一个模式的性别来作为用户的性别。此外,语音控制部 109D 根据所设置的信息(对话中介的性别信息)来识别答复的性别。一旦语音控制部 109D 以上述方式识别出用户的性别以及答复的语音的性别,则其选择各指标表中与所识别的性别的组合相对应的一个指标表。在选择指标表之后,本实施例中规则固定时段和评估时段以与第十一实施例中相同的方式交替重复。

[0350] 根据第十三实施例,使用各指标表中与用户的话语的场景相对应的一个指标表,并且还在规则固定时段中对答复的词尾的音高进行控制以相对于话语的词尾的音高呈现出由在所述一个指标表中设置的音高规则所指定的预定关系。此外,在评估时段中设置指标表的各音高规则中的实现了最逼真对话的一个音高规则。由此,第十三实施例可以容易地使得在不同场景中对话对用户而言逼真和舒服。

[0351] 第十一实施例(其中规则固定时段和评估时段重复)也可以趋向于不管场景的变化都能够容易地使得对话对用户而言舒服和逼真的条件。但是,预期到的是实现这种趋向所需的时间(即,实现这种趋向所需的规则固定时段和评估时段的重复数量)将不期望地增加。相反,如果将适当的音高规则预先设置为每个场景的初始状态,则第十三实施例可以显著减少趋向于能够容易地使对话舒服的条件所需的时间。

[0352] 虽然以上关于将如第十一实施例中使用的音高规则用作指标表的情况对第十三实施例进行了描述,但是也可以用如下方式将第十二实施例的输出规则与音高规则相结合地用于第十三实施例中:响应于场景的改变而在输出规则之间进行切换。此外,所述场景可以包括性别和年龄的组合而非仅包括年龄。此外,可以将话语的速度、答复的速度和语音合成设备10的应用(比如在类似于博物馆、美术馆和动物园的各种设施中的语音指导应用以及语音对话应用等)假定并准备为场景,而不是将所述场景限制为用户以及对用户进行答复的虚拟人物的性别和年龄。

[0353] <利用语音波形数据的修改例>

[0354] 类似于上述其他实施例,第十一至第十三实施例可以以类似于上述第二实施例的方式进行修改。即,图41的答复创建部110和语音合成部112的组合(即,利用语音合成技术的语音产生部)可以被替换为如图8或图16或图19所示的答复语音输出部113和后处理部114的组合(即,利用波形数据的语音产生部),并且可以从答复语音输出部113以例如WAV格式输出语音波形数据。

[0355] 注意,关于第一和第二实施例的各种应用示例和/或修改例也可以应用于本发明的第十一至第十三实施例。

[0356] 简言之,上述第十一至第十三实施例的特征在于:语音产生部(即,语音合成部112,或答复语音输出部113和后处理部114的组合)被构造为根据给定规则将第二区间的音高与所分析的第一区间的音高相关联,并且这些实施例还包括控制部(语音控制部109D),其在答复的发声或音响产生之后基于经由语音输入部(102)对另一话语的语音的接收而对规则进行更新。

[0357] <第十四实施例>

[0358] 下面参照图48和图49描述根据本发明的第十四实施例构造的编码/解码装置和语音合成系统。第十四实施例的特征在于通过编码/解码装置200的快速处理来高效地产生响应于输入话语的非正式反馈的语音,同时通过计算机160的高容量集中式处理来高效地产生响应于输出话语的具有相对复杂的语言含义的答复的语音。以这种方式,第十四实施例提供了能够实现灵活人机对话功能的语音合成系统(语音合成设备100)。

[0359] 图48是示出根据本发明的第十四实施例构造的包括编码/解码装置的语音合成设备100(语音合成系统)的硬件配置的图。该语音合成设备100(语音合成系统)是例如比如便携式电话的终端设备的形式,其包括:语音输入部102;扬声器142;编码/解码装置(下文称作“CODEC”(编码器解码器))200;以及主机计算机160,其相对于CODEC 200是更高级别的计算机。

[0360] 单芯片或多芯片模块形式的CODEC 200包括微处理器202、存储器204、A/D转换器206和D/A转换器208。此外,CODEC 200被构造为通过微处理器202执行存储在存储器204中的程序P1(固件)来对语音信号进行处理。

[0361] 主机计算机160包括CPU(中央处理单元)162和存储器164。CPU 162经由总线BUS与CODEC 200和存储器164相连接。在本实施例中用于语音处理的程序P2以及操作系统存储在存储器164中。本实施例中的语音处理包括:当用户讲出话语时用于利用合成的语音输出针对该话语的答复或非正式反馈的对话处理;用于将用户的话语转换为文字列的语音识别处理;以及用于再现和输出通过主机160处理的音乐、语音等的再现(播放)处理。

[0362] 虽然没有特别示出,但是语音合成设备100还包括显示部、操作输入部等,以使得用户能够检查语音合成设备100的状态以及向设备100输入各种操作。此外,语音合成设备100可以是笔记本或平板个人计算机的形式而不是比如便携式电话的终端设备的形式。

[0363] 图49是示出本发明的语音合成设备100的功能配置的功能性框图。通过CODEC 200中的执行程序P1的微处理器202以及通过主机 160中的执行程序P2的CPU 162来建立各个功能块。如图中所示, CODEC 200内已建立有音高分析部106、语言分析部108a、协作部140、读出部142、非正式反馈数据存储部143和音高控制部144,而主机 160内已建立有语言分析部108b、协作部180、答复创建部110、语言数据库122、答复数据库124和语音库128。用与图1中相同的参考标记示出与图1中的块实现基本相同功能的图49中的块,并且将不对其进行描述以避免不必要的重复。

[0364] CODEC 200中的音高分析部106和语言分析部108a与图1中示出的音高分析部106和语言分析部108以基本相同的方式工作。但是,注意,语言分析部108a确定经由语音输入部102接收到的话语是否具有应当对其返回非正式反馈的内容或者是否具有应当对其返回不同于非正式反馈的答复的内容。语言分析部108a进行这种确定的具体方式的可能示例之中的一个示例是:预先存储朝向例如典型询问(即,针对其要返回不同于非正式反馈的答复的话语)的词尾的音量和音高变化模式。在这种情况下,如果由从音高分析部106输出的音量波形和音高波形表示的音量变化和音高变化匹配于(或高度类似于)任意的预先存储的音量变化模式和音高变化模式,则该话语被确定为询问。另一方面,如果由从音高分析部106输出的音量波形和音高波形表示的音量变化和音高变化不匹配于(或低度类似于)任意的预先存储的音量变化模式和音高变化模式,则该话语被确定为应当针对其返回非正式反馈的话语。注意,如果语言分析部108a无法确定是应当返回非正式反馈还是应当返回不同于非正式反馈的另一答复,或者如果语言分析部108a的确定结果可靠度低,则可以采用这样的特定规则:其使得确定步骤被直接委托给在随后阶段设置的主机160(语言分析部108b)。此外,由于非正式反馈旨在单纯地使交谈流畅(或者改善交谈的节奏)并且可以无视非正式反馈的含义,实际上即使在话语是询问的情况下返回非正式反馈问题也不大。

[0365] 如果确定经由语音输入部102接收到的话语具有应当对其返回非正式反馈的内容,则协作部140控制音高控制部144选择非正式反馈数据作为处理对象并通知主机160其无需对该话语执行处理。在确定经由语音输入部102接收到的话语具有应当对其返回不同于非正式反馈的另一答复的内容的情况下(在应用特定规则的条件下,包括话语的内容无法辨识的情况和确定的结果可靠度低的情况),协作部 140指示音高控制部144选择由主机160的答复创建部110所创建的答复数据(即,不同于非正式反馈数据的响应数据)作为音高控制处理的对象,并通知主机160其应当创建(或获取)针对话语的答复(响应)。

[0366] 非正式反馈数据存储部143中存储有非正式反馈数据的多个集合。这里,非正式反馈数据的多个集合是语音波形数据的多个集合,该语音波形数据包括简单答复,比如“êto”(“让我想想。”)、“naru hodo”(“我知道了。”)、“sou desune”(“恩,……。”)、“a hai”(“哦,是的。”)、“nn”(“让我们看看。”)。语音波形数据的集合例如为WAV格式。

[0367] 如果语音输入部102确定经由语音输入部102接收到的话语具有应当对其返回非正式反馈的内容,则读出部142根据预定顺序或随机读出非正式反馈数据的多个集合中的任何一个,并将所读出的非正式反馈数据供应至音高控制部144。通过这种非正式反馈数据

独立于话语的具体内容而输出的构造,有时可以输出与话语无关的非正式反馈,但是,这并不重要,因为可以如上所述地无视非正式反馈的含义。

[0368] 因为由读出部142读出的非正式反馈和由答复创建部110供应的答复数据(不同于非正式反馈的响应数据)都是语音波形数据,因此可以通过按原样再现的非正式反馈和响应数据以语音输出非正式反馈和答复(不同于非正式反馈的响应)。但是,第十四实施例被构造为在考虑要针对其返回非正式反馈或答复的话语的音高的情况下对非正式反馈或答复(响应)的音高进行控制,如上述第一实施例等中描述的那样,并且这样的音高控制功能通过音高控制部144来执行。即,与上述第一实施例等一样,音高控制部144根据音高分析部106所分析的话语的第一区间(例如,词尾)的音高来控制基于从非正式反馈数据存储器143读取的非正式反馈数据或从答复创建部110 供应的答复数据的语音的各音高。更具体地,音高控制部144通过控制非正式反馈或答复的整个答复语音的各音高使得非正式反馈数据的特定区间(第二区间)或答复数据的特定区间(第三区间)的音高与话语的第一区间的音高呈现出预定的音程关系,从而控制答复语音合成。由音高控制部144合成的答复语音信号被D/A转换器208转换为模拟信号并随后经由扬声器142可听地输出。

[0369] 另一方面,在主机160中,协作部180将诸如各种参数和各种状态之类的信息供应至CODEC 200的协作部140。此外,当从协作部 140接收到创建答复(响应)的请求时,协作部180控制语言分析部 108b和答复创建部110。与图1的语言分析部108一样,语言分析部108b对被A/D转换器206转换成数字信号的语音信号所表示的话语的含义进行分析。在图49示出的主机160中,语言分析部108b和答复创建部110起到与图1示出的语言分析部108和答复创建部110 基本相似的作用。注意,图49的主机160中的答复创建部110原则上被构造为创建不同于非正式反馈的答复(响应)。但是,如果CODEC 200中的语言分析部108a无法辨识话语的内容,或者如果语言分析部108a的辨识结果的可靠度低,则主机160中的答复创建部110有时创建非正式反馈作为针对话语的答复。由答复创建部110创建或获取的答复数据被供应至音高控制部144。

[0370] 下面描述语音合成设备100的行为。图50是在语音合成设备100 中执行的语音处理的流程图。首先,响应于用户执行预定的操作(例如,在主菜单(未示出)上选择与语音处理相对应的图标),CPU 162 启动程序P2,从而在CODEC 200和主机160中建立图49示出的各功能块。

[0371] 在步骤S11,一旦用户输入话语的语音至语音输入部102,则该语音被语音输入部102转换为语音信号并随后利用A/D转换器206 转换为数字信号。随后,在步骤S12,主机160中的协作部180确定在已开始的语音处理中是否指定了执行对话处理。如果在已开始的语音处理中未指定执行对话处理(步骤S12判定为否),则在步骤S16 执行其他处理。其他处理的示例包括:语音辨识处理,其用于将被转换为数字信号的语音信号供应至主机160中的如图49中的标记※1 所指示的另一功能块(未示出),使得其他功能块将用户的话语转换为文字列;以及再现处理,其利用D/A转换器208将其他块处理的数据转换为模拟信号并通过扬声器142可听地再现该模拟信号。

[0372] 另一方面,如果在已开始的语音处理中指定了执行对话处理(步骤S12判定为是),则协作部180通知CODEC 200的协作部140该结果,使得协作部140指示语言分析部108a在步骤S13确定输入话语是否具有应当对其返回非正式反馈的内容。如果输入话语具有应当对

其返回非正式反馈的内容(步骤S13判定为是),则在步骤S14执行如下非正式反馈处理。另一方面,如果输入话语不具有应当对其返回非正式反馈的内容(步骤S13判定为否),则在步骤S15执行如下答复处理。注意,该语音处理在步骤S14、S15和S16之后结束。

[0373] 图51是示出在步骤S14执行的非正式反馈处理的细节的流程图。首先,已确定输入话语具有应当对其返回非正式反馈的内容的语言分析部108a通知协作部140该结果,并且协作部140通知主机160的协作部180无需创建针对该话语的答复(步骤Sa41)。在从协作部140接收到这种信息时,协作部180指示语言分析部108b忽略与该话语相对应的数字信号。由此,在主机160中不执行关于该话语的语音处理(步骤Sb31)。

[0374] 同时,在步骤Sa42,音高分析部106例如用以下方式来分析输入话语的语音信号从而对该话语的第一区间(例如,词尾)的音高进行识别并随后将指示识别出的音高的音高数据供应至音高控制部144。与该音高分析并行,在步骤Sa43,读出部142选择性地读出存储在非正式反馈数据存储部143中的非正式反馈数据的集合中的任一个并将所读出的非正式反馈数据供应至音高控制部144。与上述实施例一样,音高控制部144对与第二区间(例如,词尾)相对应的一部分简单再现的非正式反馈数据的音高进行分析,随后以如下方式将整个非正式反馈数据的各音高进行移动:使所分析的音高与由从音高分析部106供应的音高数据所指示的音高(诸如词尾之类的第一区间的音高)呈现出预定的音程(例如,向下五度)关系,并随后输出音高移动后的非正式反馈数据(步骤Sa44)。音高控制部144对音高移动后的非正式反馈数据进行再现并将再现的音高移动后的非正式反馈数据供应至D/A转换器208。以这种方式,输出与话语相对应地移动了音高的非正式反馈。在输出音高移动后的非正式反馈数据之后,该语音处理(图50)与非正式反馈处理一同结束。

[0375] 下面描述该语音处理的步骤S15处执行的答复处理。图52是示出该答复处理的细节的流程图。已确定输入话语具有不应当对其返回非正式反馈的内容的语言分析部108a通知协作部140该结果,使得协作部140通知主机160的协作部180应当创建针对该话语的答复(答复请求)(步骤Sa45)。在接收到该通知时,协作部180指示语言分析部108b对话语的含义进行分析。随后,语言分析部108b在步骤Sb32按指示对该话语的含义进行分析。随后,在步骤Sb33,答复创建部110创建(获取)与该话语的含义相对应的答复数据并将该答复数据供应至音高控制部144。

[0376] 同时,在答复处理中,在步骤Sa46,以与上述步骤Sa42相同的方式,音高分析部106识别输入话语的第一区间(例如,词尾)的音高并随后将指示识别出的音高的音高数据供应至音高控制部144。随后,音高控制部144在答复数据简单再现时对该答复数据的第三区间(例如,词尾)的音高进行分析,而且还将整个答复数据的各音高进行移动使得所分析的音高与从音高分析部106供应的音高数据所指示的音高呈现出向下五度关系,随后音高控制部144将音高移动后的答复数据输出(步骤Sa47)。

[0377] 通过语音合成设备100的这样的实施例,无论是利用非正式反馈数据输出非正式反馈还是利用答复数据输出不同于非正式反馈的另一答复,可以给用户好似正在进行人与人之间的对话的自然感觉。此外,根据本实施例,当针对话语要返回非正式反馈时,仅在CODEC 200中处理非正式反馈数据而无需在主机160中进行处理,由此,可以响应良好地快速输出非正式反馈。此外,当响应于话语要输出不同于非正式反馈的另一答复时,在主机160中创建或获取这样的答复,从而可以提升答复的精确度。因此,本实施例不仅允许响应

良好地快速输出非正式反馈,而且允许以更高的精确度输出不同于非正式反馈的另一答复。

[0378] <话语和非正式反馈>

[0379] 第十四实施例被描述为按如下方式构造:当确定话语具有应当对其返回非正式反馈数据的内容时,读出在非正式反馈数据存储部 143中存储的非正式反馈数据的集合中的任一个。然而,通过这种构造,虽然如上所述并不重要,但是有时会输出与话语无关的非正式反馈。因此,可以提前将非正式反馈数据的集合与同非正式反馈相对应的可能话语的典型音量变化模式和音高变化模式相关联,如此,当语言分析部108a确定了从音高分析部106供应的话语的音量和音高变化匹配于任一话语的音量和音高变化模式时,其指示读出部142读出与该话语相关联的非正式反馈数据的集合。更具体地,通过这样的构造,如果非正式反馈数据的集合“sou desune”(“是啊。”)与话语“samui nâ”(“很冷,是吧?”)的音量和音高变化相关联地预先存储在非正式反馈数据存储部143中,并且如果用户的实际话语“samui nâ”(“很冷,是吧?”)的音量和音高变化匹配所存储的话语“samui nâ”的音量和音高变化,则可以输出恰当的非正式反馈“是啊。”因此,在这种情况下,可以避免输出不恰当的或无关的非正式反馈比如“êto”(“让我想想。”)、“naru hodo”(“我知道了。”)或“sorede”(“所以呢?”)

[0380] <语言分析部>

[0381] 虽然以上关于CODEC 200包括语言分析部108a并且主机160包括语言分析部108b的情况对第十四实施例进行了描述,但是可以省略任一语言分析部,比如语言分析部108a。如果省略了语言分析部 108a,则语言分析部108b执行语言分析部108a的功能。即,语言分析部108b确定数字语音信号所表示的话语是否具有应当对其返回非正式反馈的内容。如果该话语具有应当对其返回非正式反馈的内容,则语言分析部108b可以经由协作部180和140直接或间接地将该结果的信息提供给CODEC 200的读出部142并指示读出部142读出非正式反馈数据。

[0382] 总结CODEC(编码/解码装置)200的主要构造特征,CODEC 200 包括:A/D转换器(206),其将话语的输入语音信号转换为数字信号;音高分析部(106),其基于数字信号对话语的第一区间的音高进行分析;非正式反馈获取部(语言分析部108a、读出部142和非正式反馈数据存储部143的组合),其在针对话语要返回非正式反馈时,获取与该话语的含义相对应的非正式反馈数据;音高控制部(144),其以如下方式控制所获取的非正式反馈数据的音高:使所述非正式反馈数据的第二区间具有与所分析的第一区间的音高相关联的音高;以及D/A转换器(208),其构造为将音高受控的非正式反馈数据转换成模拟信号。

[0383] 注意,关于第一实施例和第二实施例的各种应用示例和/或修改例也可以应用于本发明的第十四实施例。

[0384] 此外,主机计算机160的主要功能特点可以总结如下。当针对话语应当返回不同于非正式反馈的答复语音时,主机计算机160被构造为根据由A/D转换器(206)转换的数字信号来获取响应于该话语的答复语音数据并随后将所获取的答复语音数据返回至编码/解码装置(CODEC 200)。编码/解码装置(CODEC 200)的音高控制部(144)还被构造为对从主机电脑(160)返回的答复语音数据的音高进行控制,使得该答复语音数据的第三区间具有与所分析的第一区间的音高相关联的音高,并且D/A转换器(208)还被构造为将音高受控的答

复语音数据转换为模拟信号。

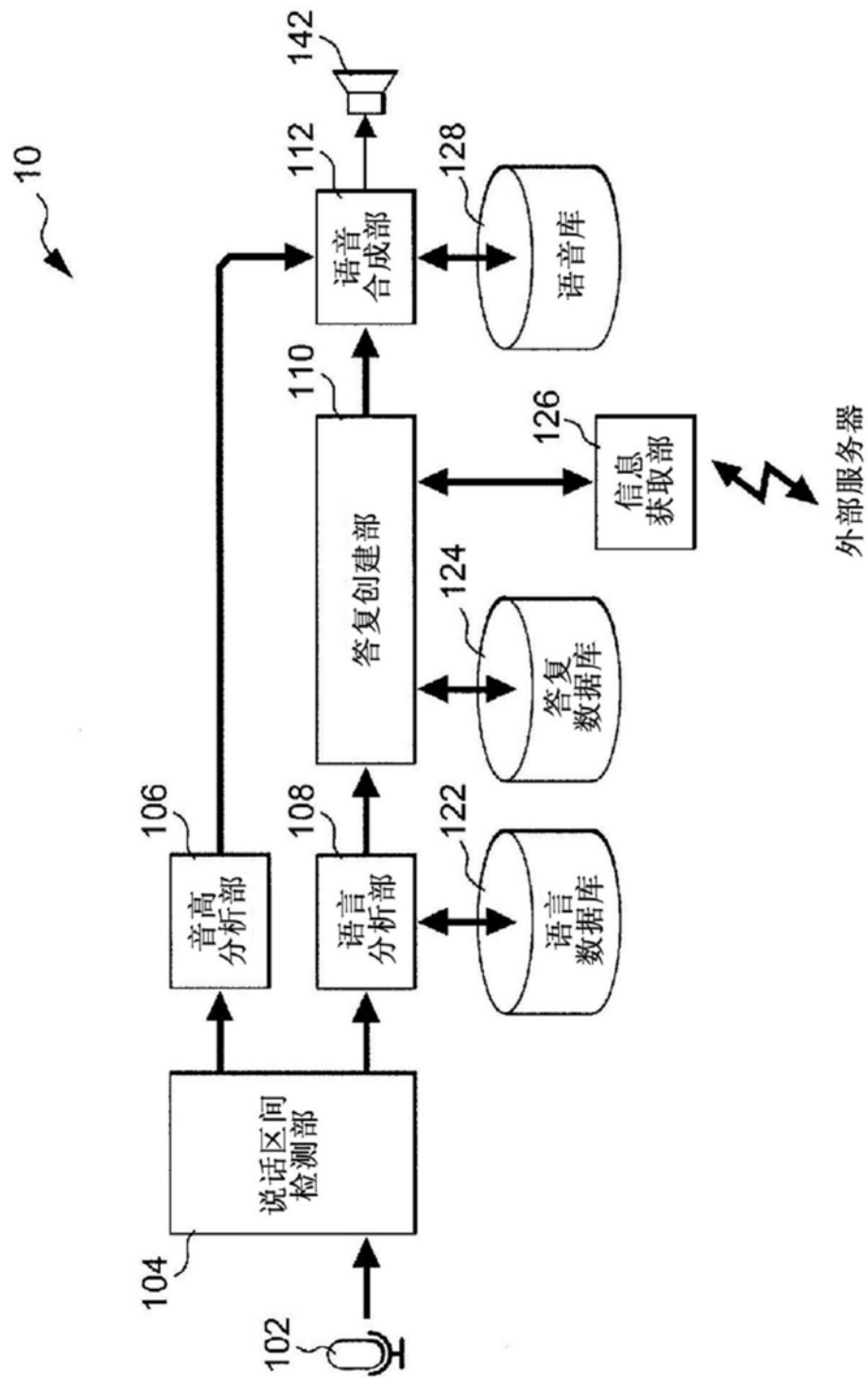


图1

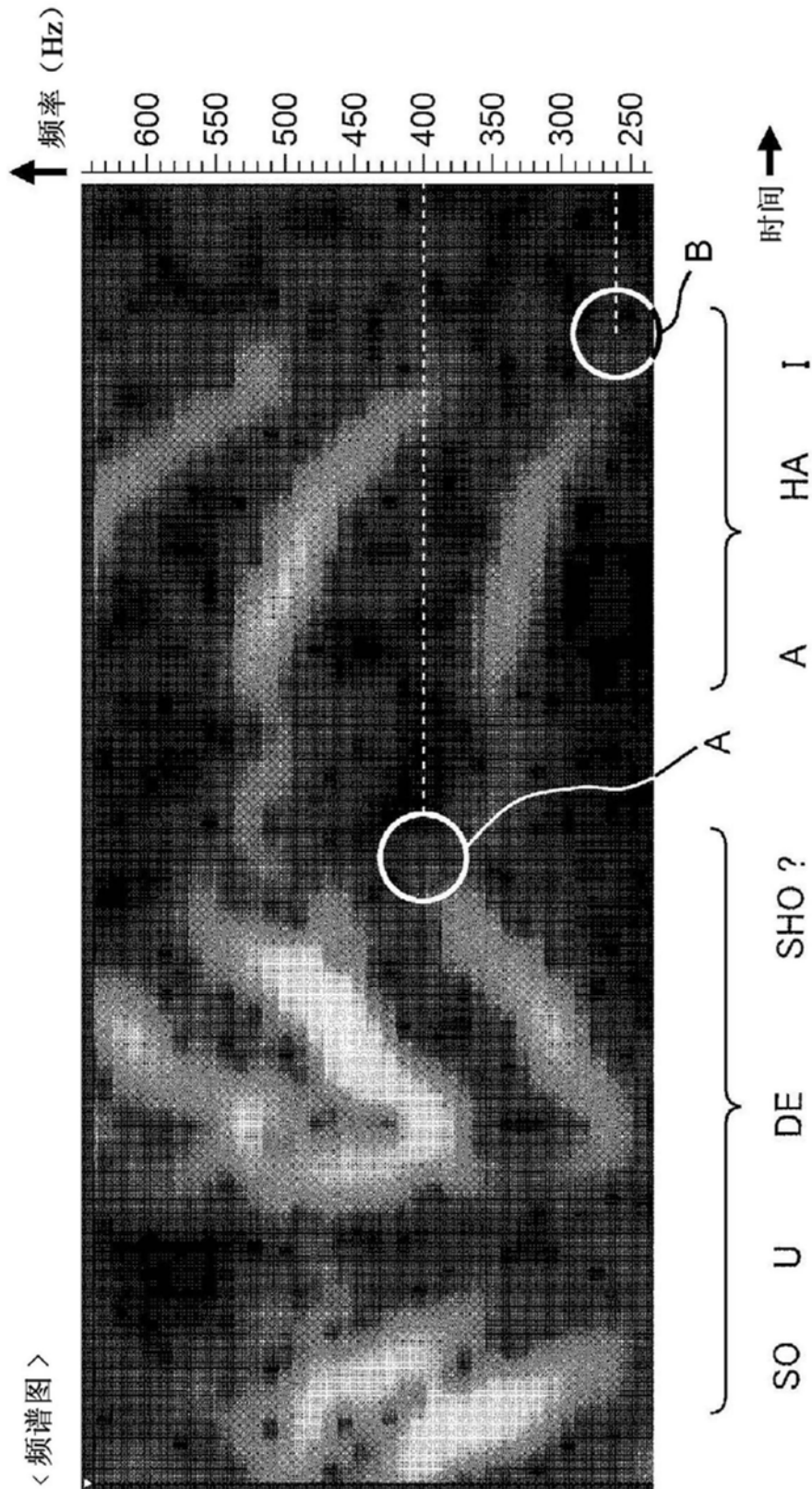


图2

＜ 音名 & 频率之间的关系 ＞

八度		3				4						5	
音名 (音节名)	...	G (SO)	A (LA)	B (SI)	C (DO)	D (RE)	E (MI)	F (FA)	G (SO)	S (LA)	B (SI)	C (DO)	...
频率 (Hz)	...	198.0	220.0	247.5	264.0	297.0	330.0	352.0	396.0	440.0	495.0	528.0	...
基准 (C 大调)	...	3/4	5/6	15/16	1/1	9/8	5/4	4/3	3/2	5/3	15/8	2	...
基准 (A 小调)	...	9/10	1/1	9/8	6/5	4/3	3/2	8/5	9/5	2	9/4	12/5	...

图3

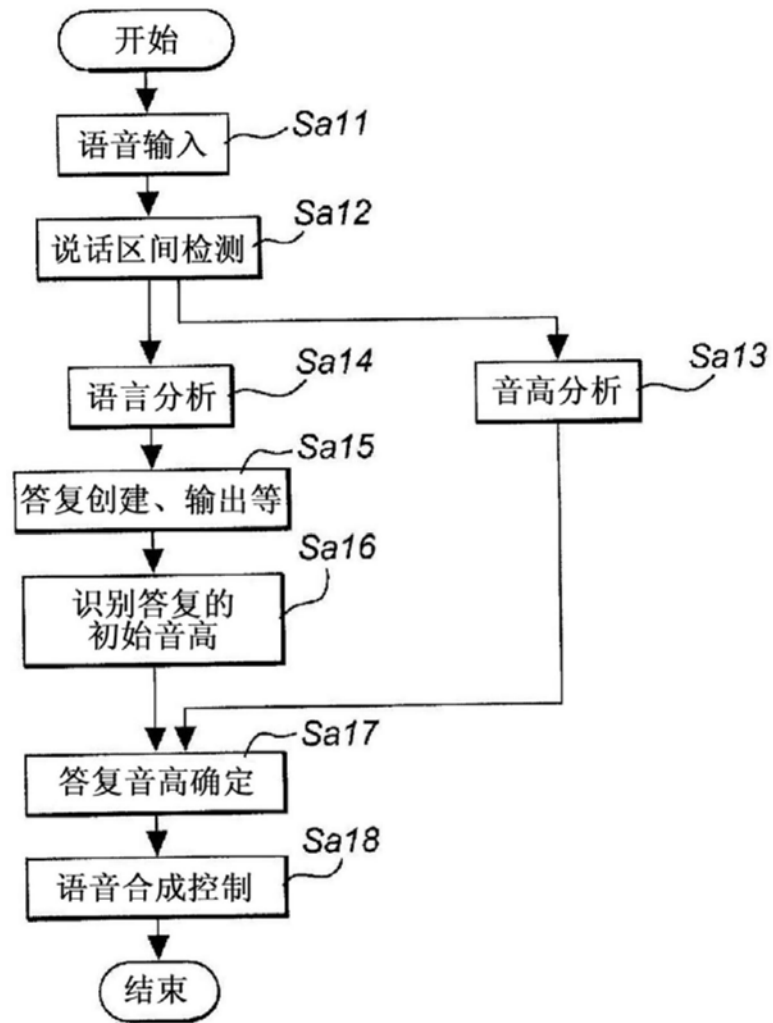


图4

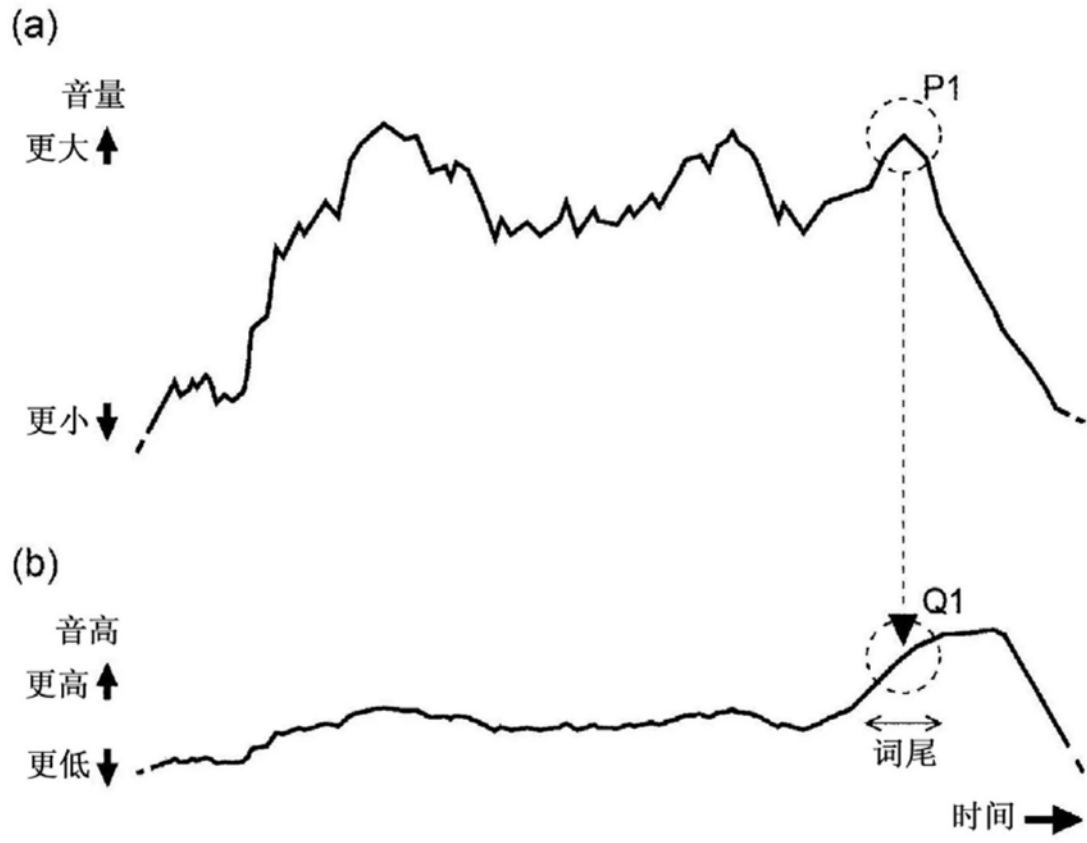


图5

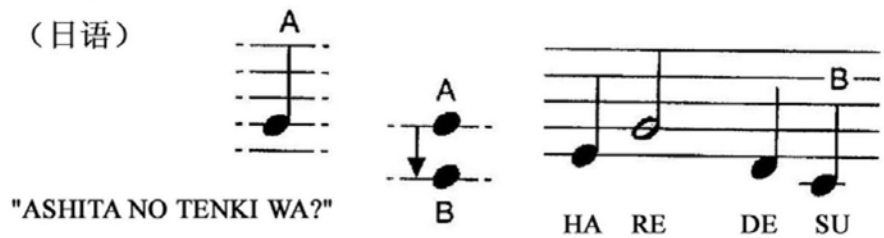
(a) < 语音序列 (基本语音) >

(日语)



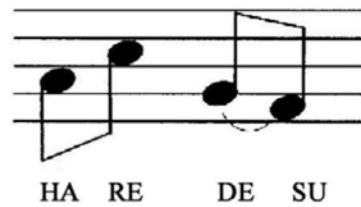
(b) < 音高移动后 >

(日语)

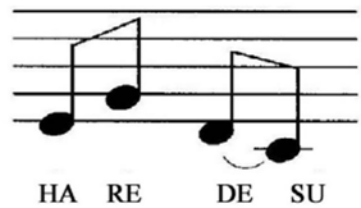


(c) < 语音序列 (基本语音) >

(日语)



(d) < 音高移动后 > (日语)



(e) < 音高移动后 > (英语)



(f) < 音高移动后 > (中文)

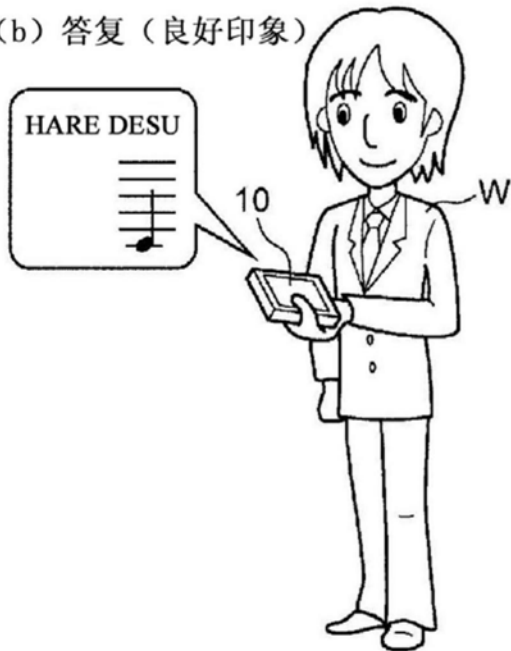


图6

(a) 询问



(b) 答复 (良好印象)



(c) 答复 (不好印象)



图7

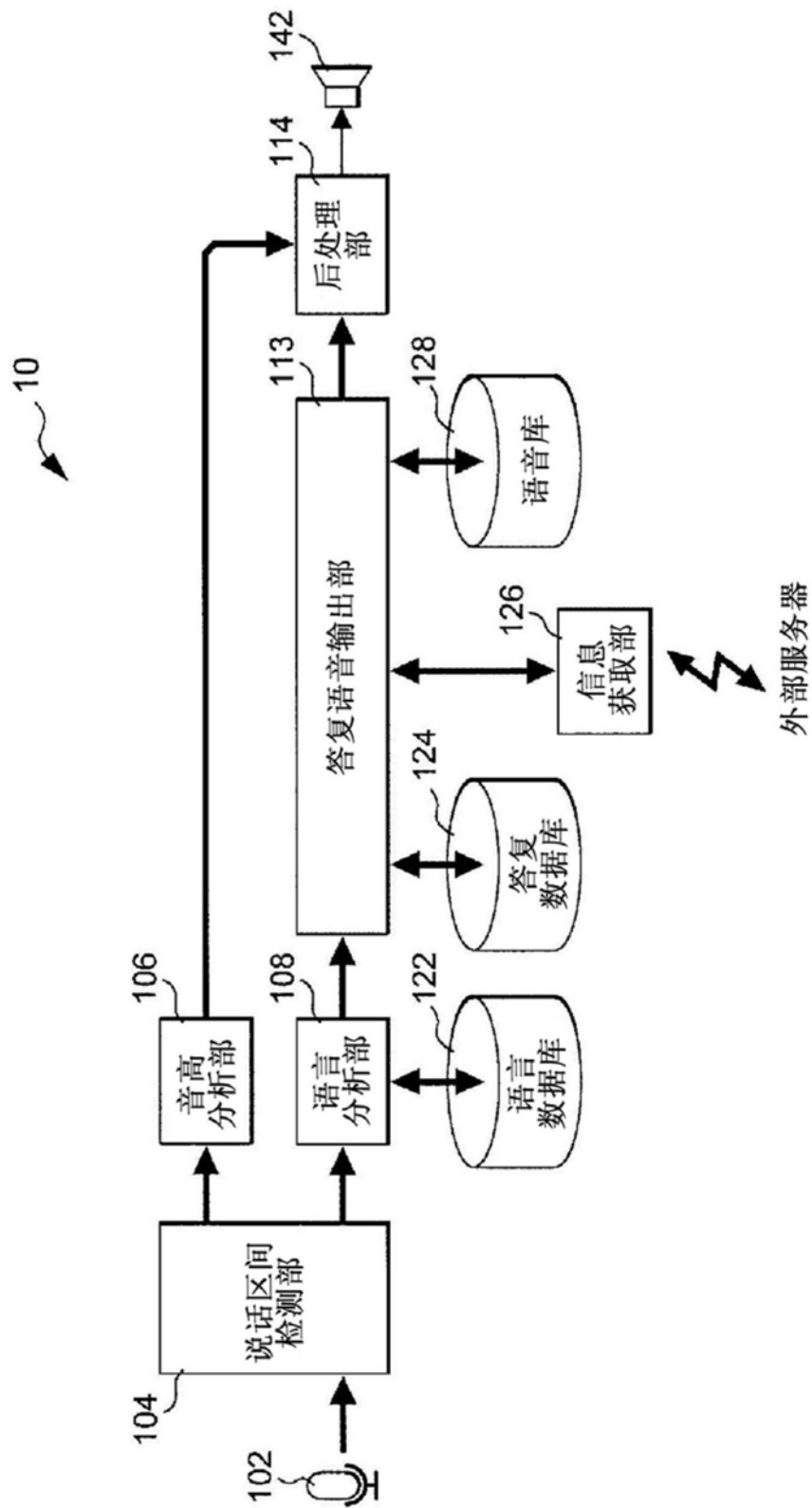
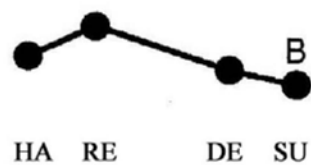


图8

(a) < 语音波形数据 >



(b) < 音高转换 >

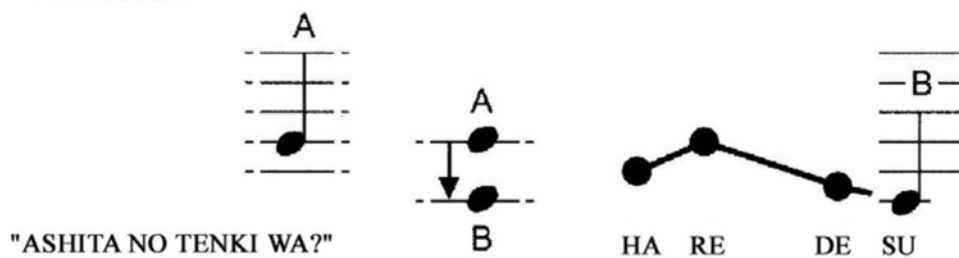


图9

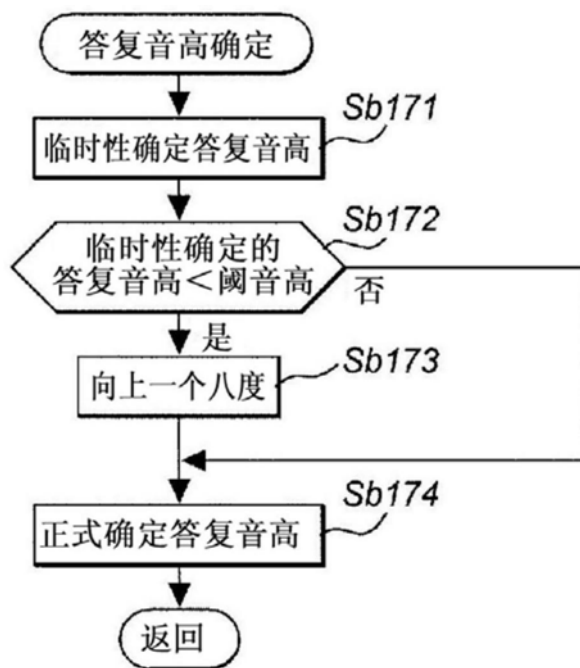


图10

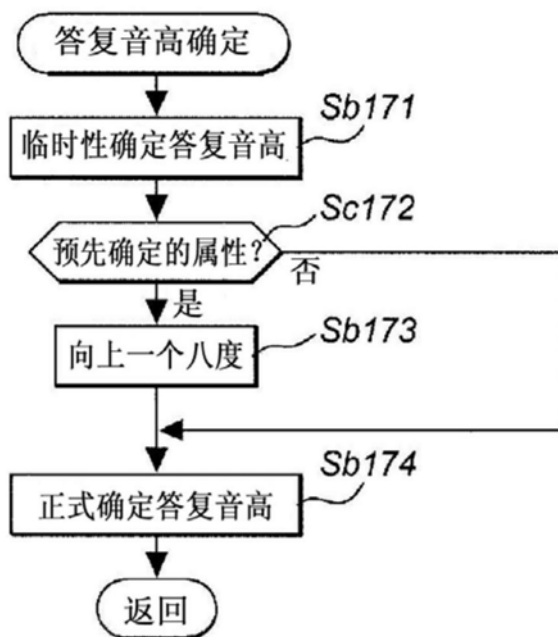


图11

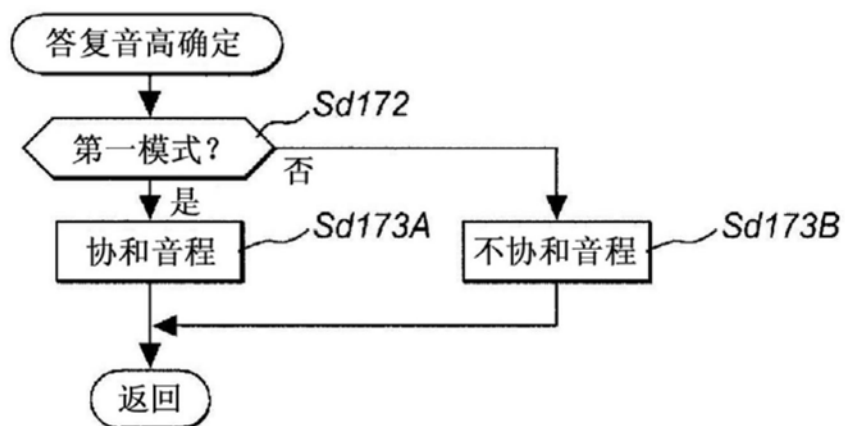
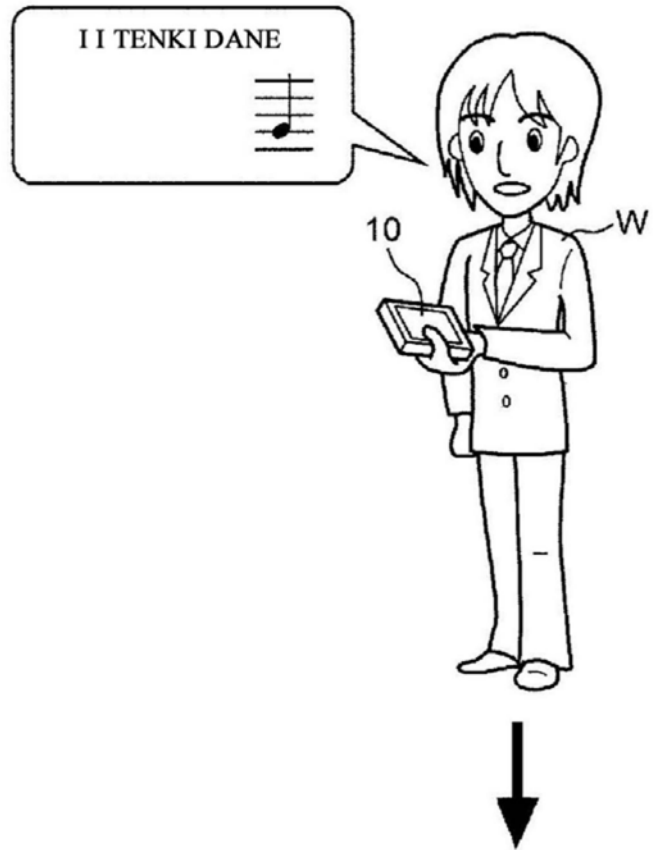


图12

(a) 询问



(b) 答复



图13

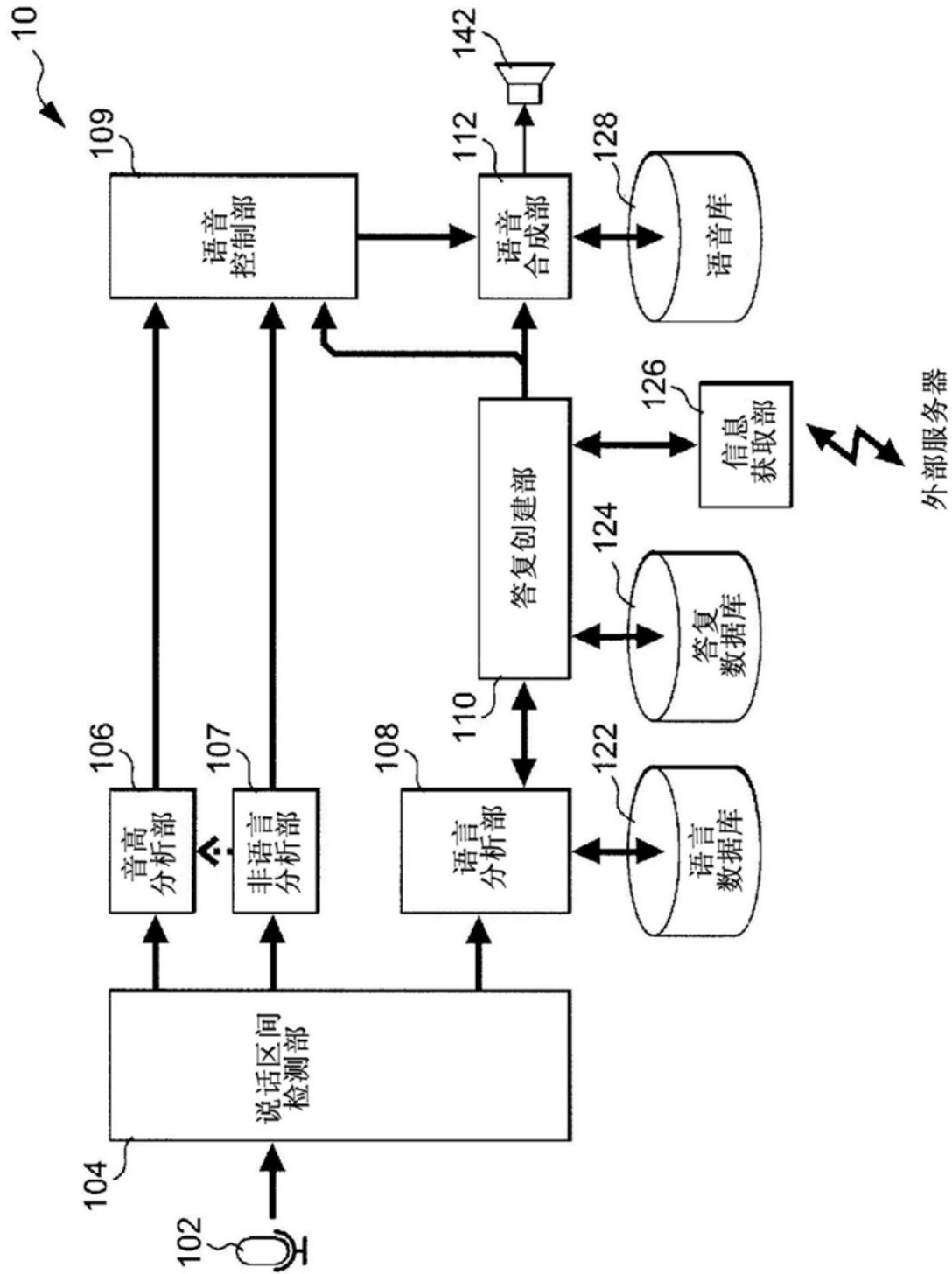


图14

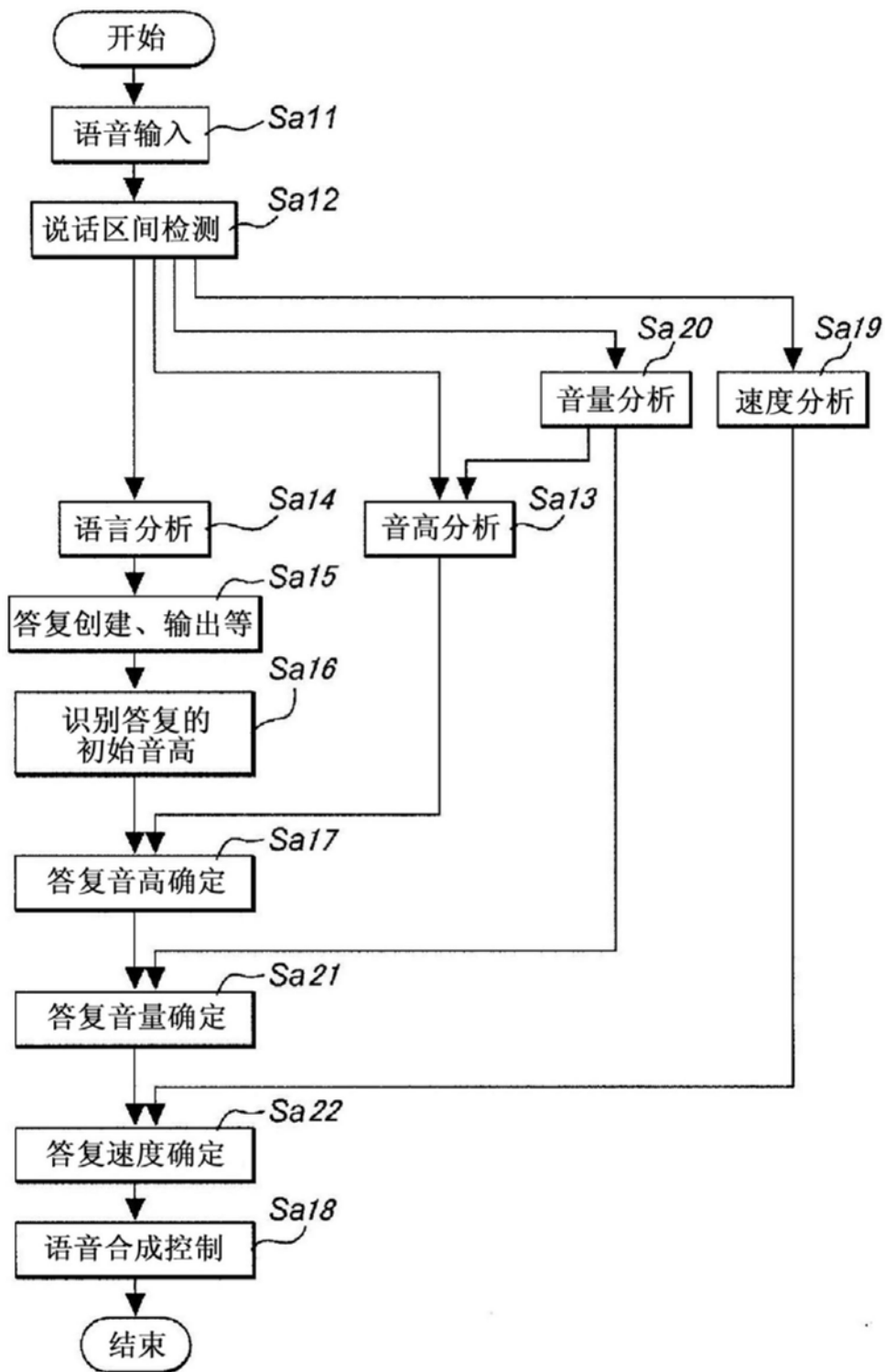


图15

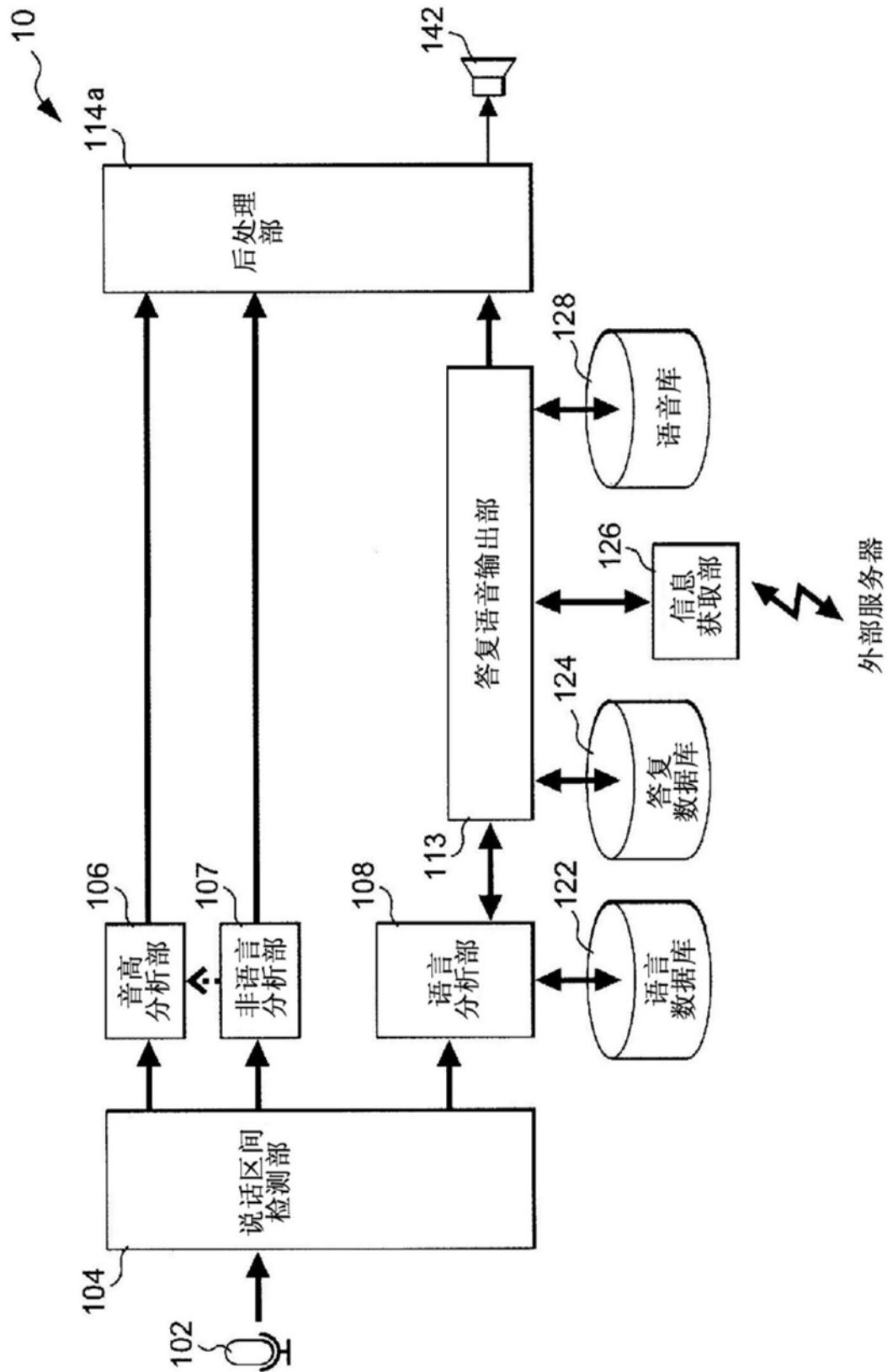


图16

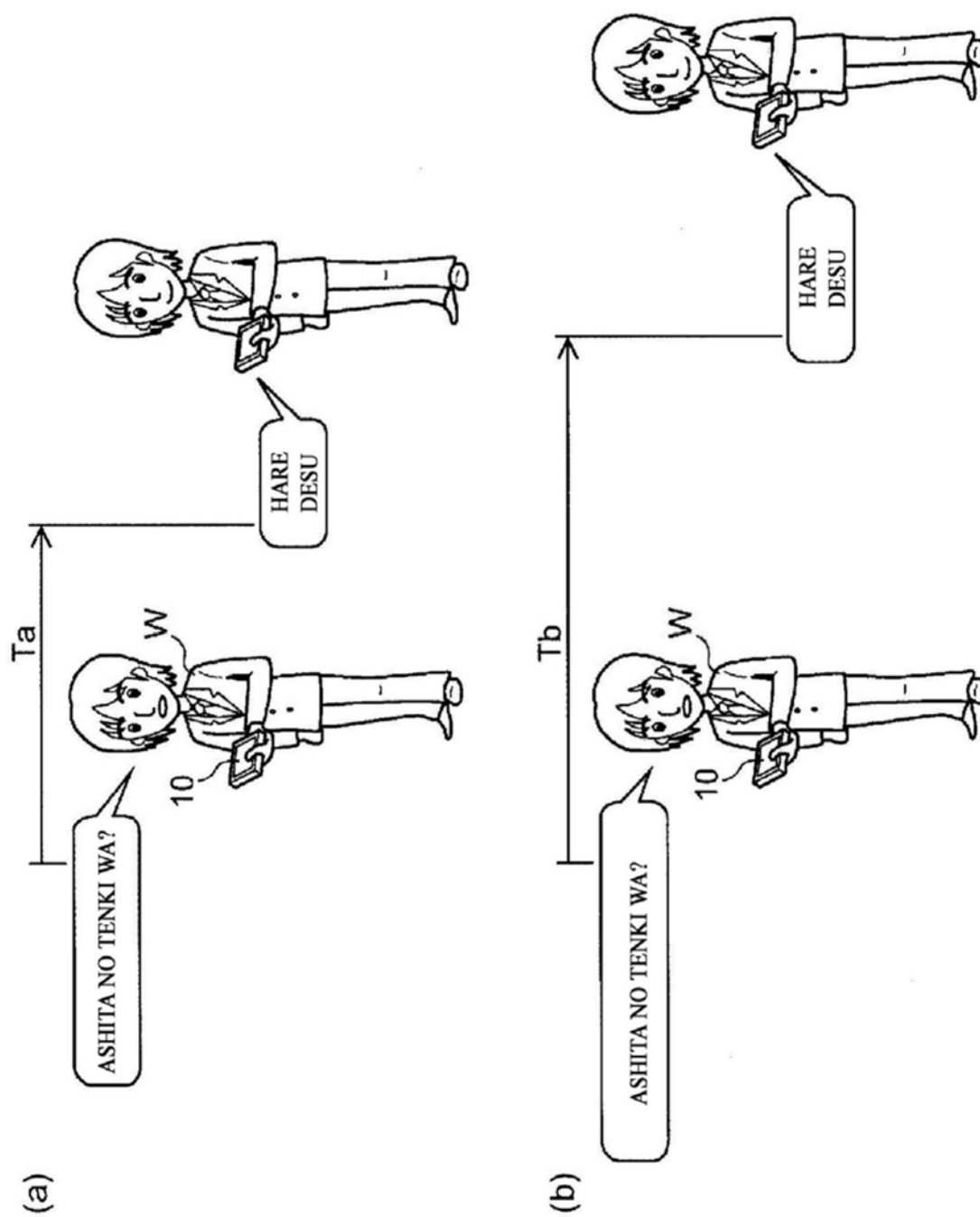


图17

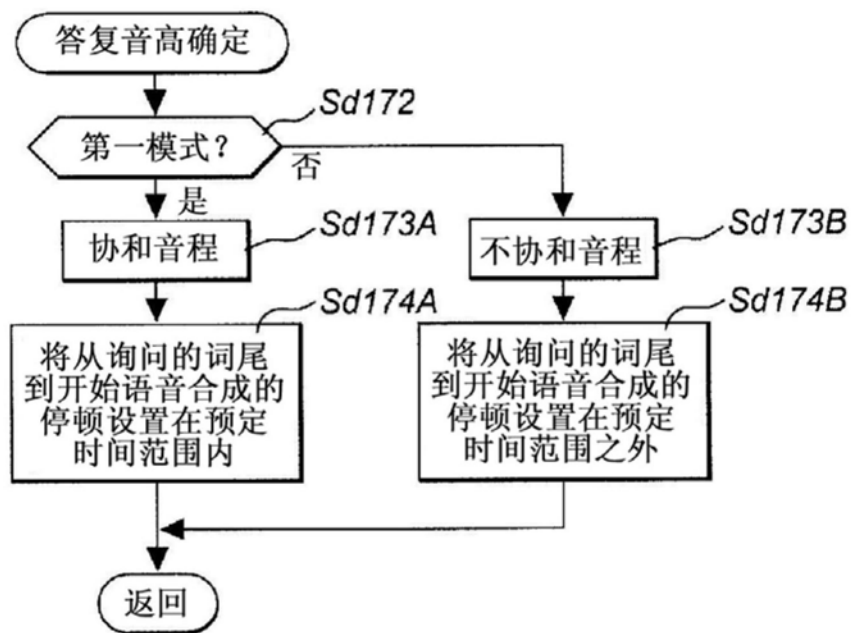


图18

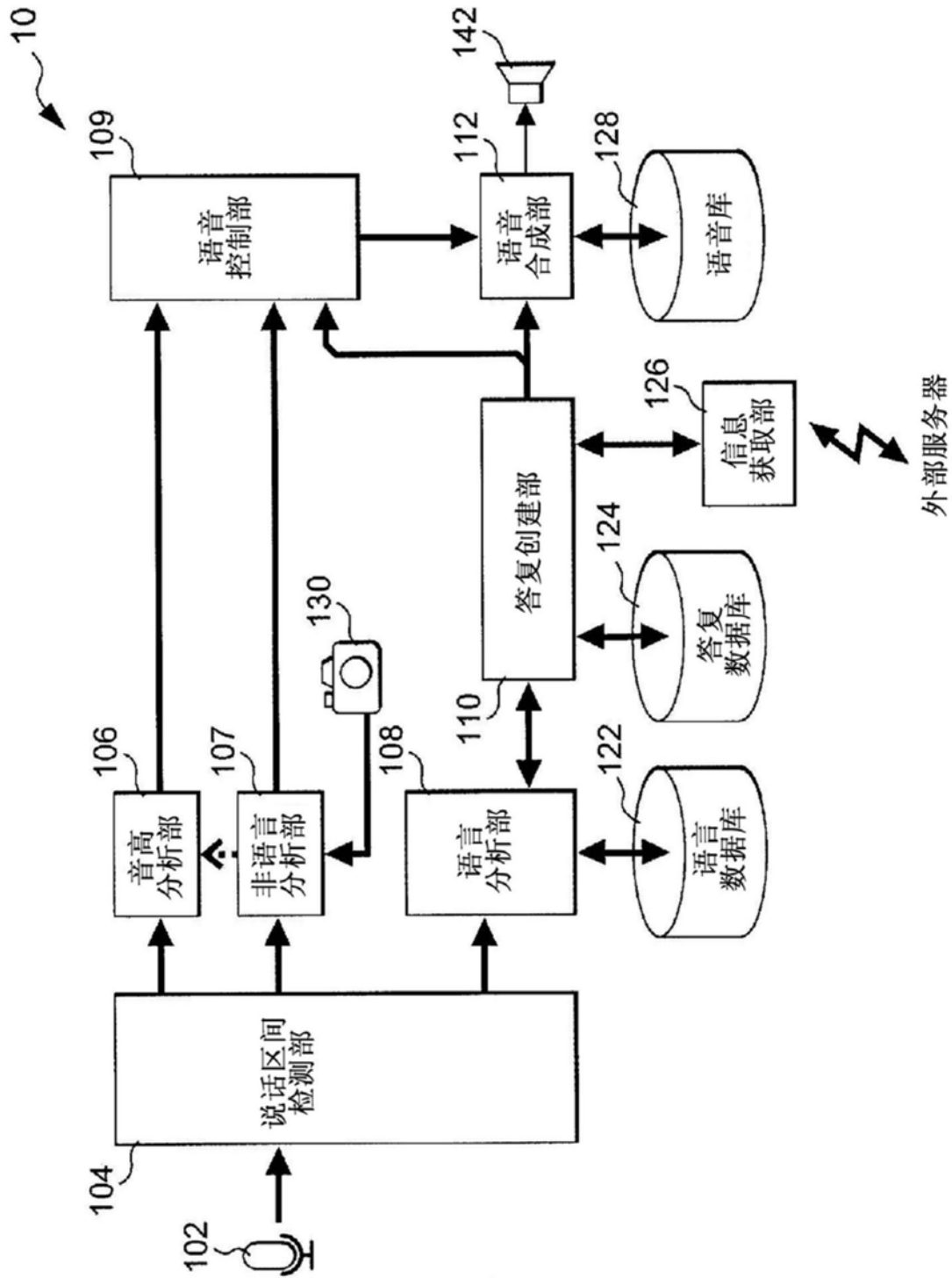


图19

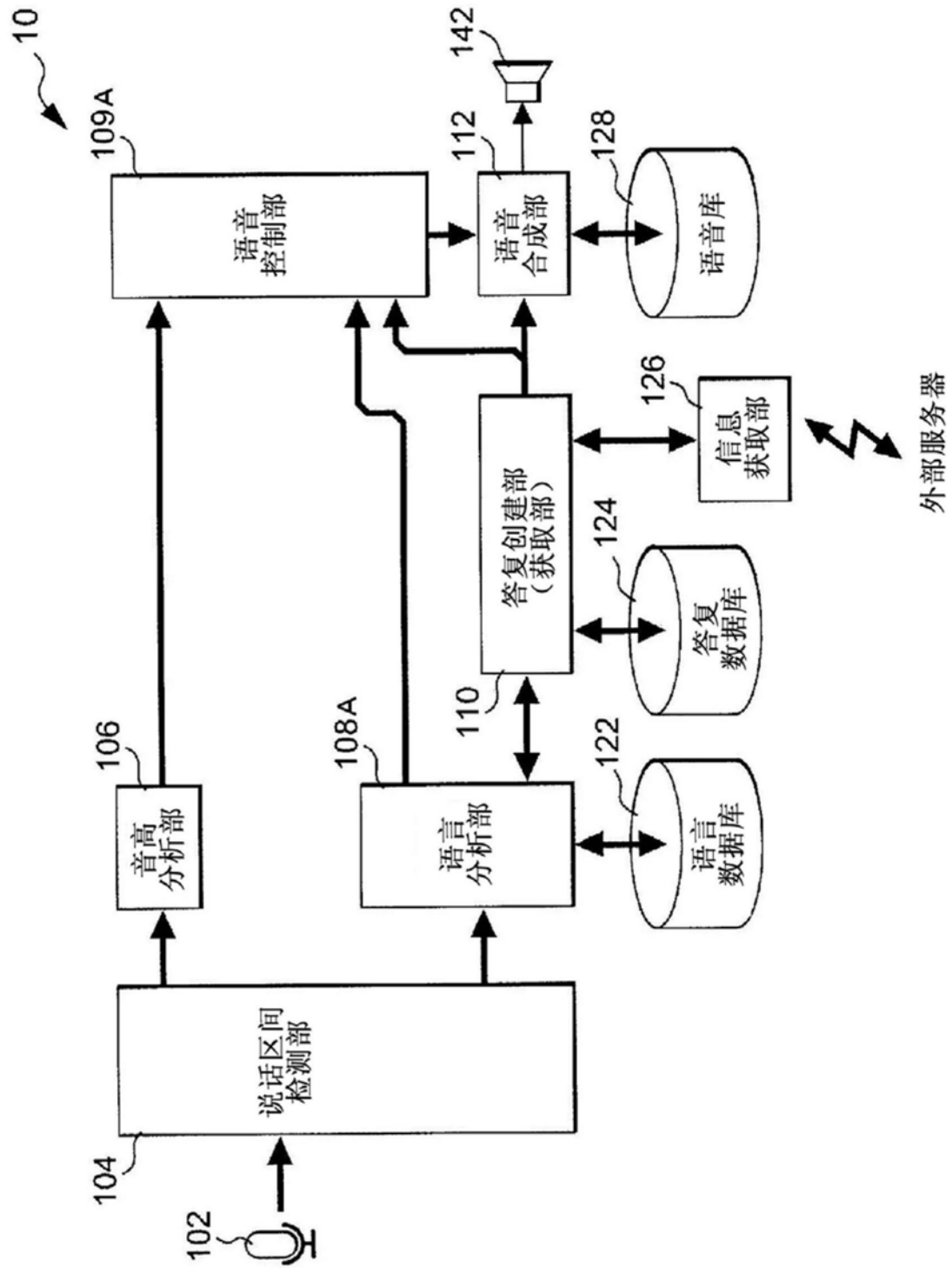


图20

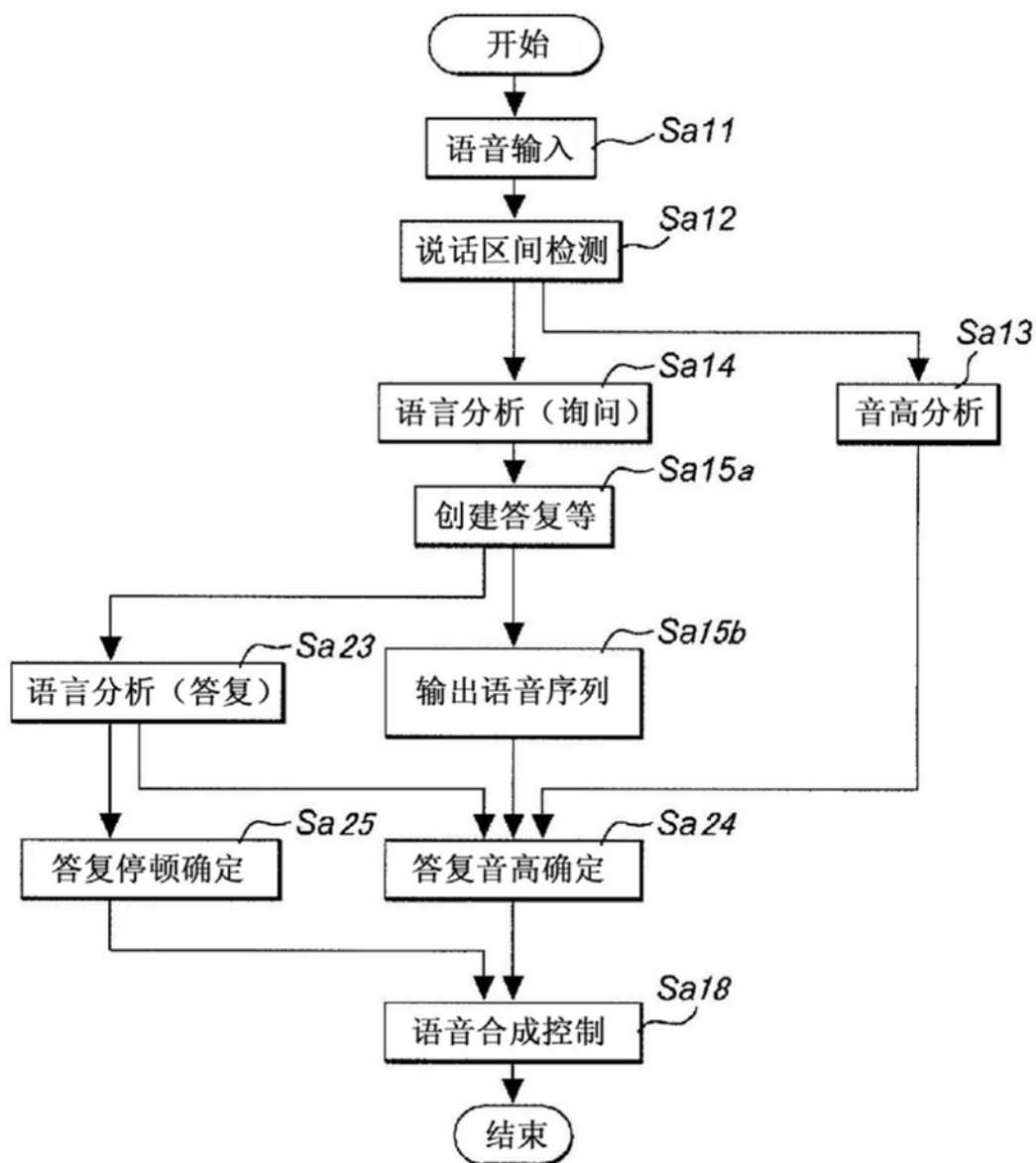
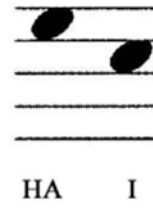


图21

(a) 语音序列



(b) 音高移动：肯定 / 同意

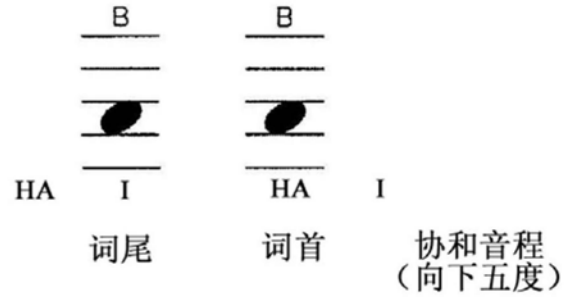
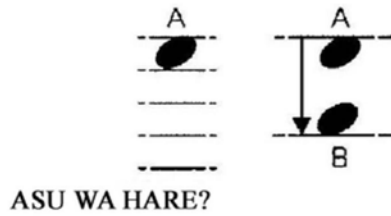


图22

(a) 语音序列



(b) 音高移动：否定 / 不同意

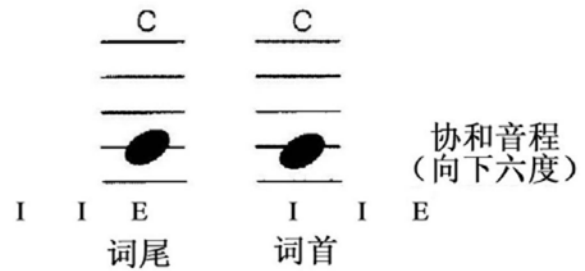
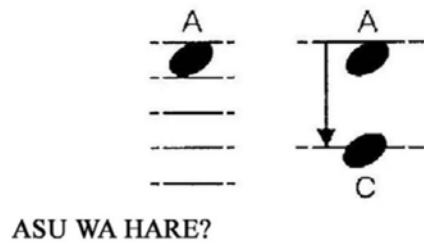
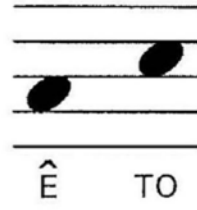


图23

(a) 语音序列



(b) 音高移动：共鸣的非正式反馈

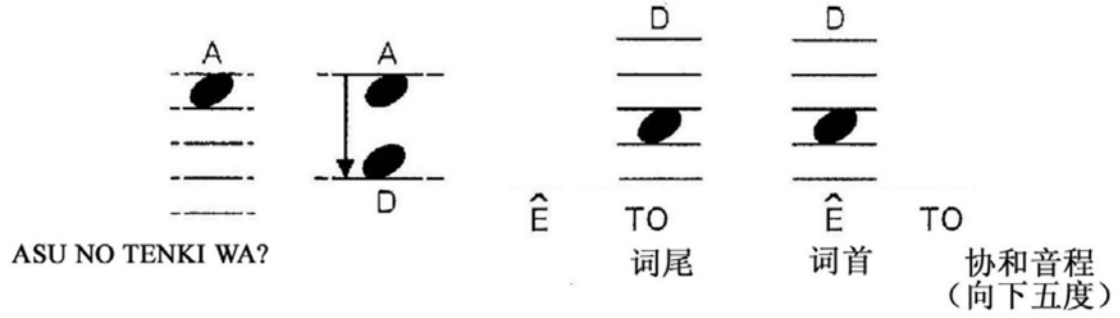


图24

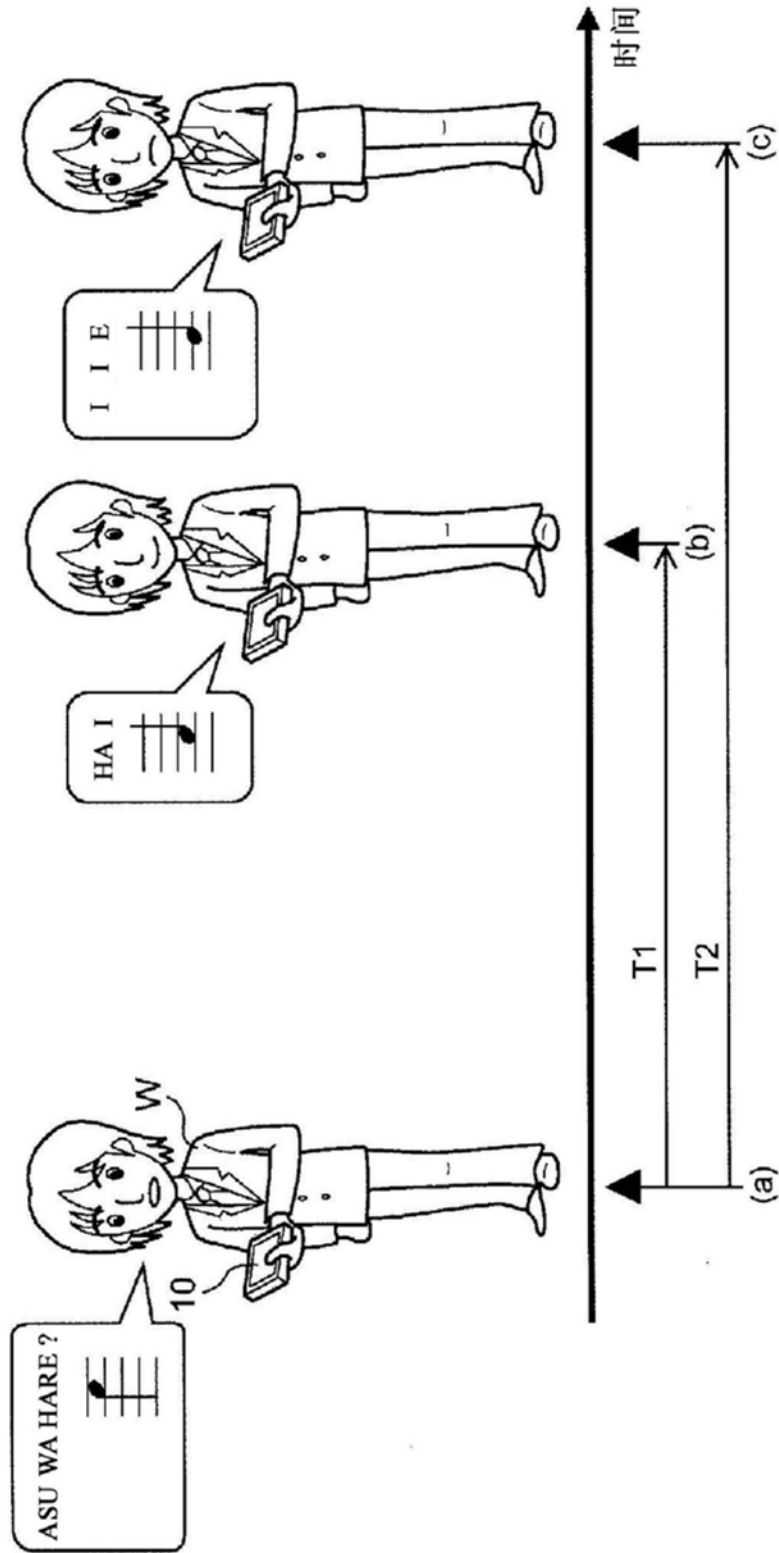


图25

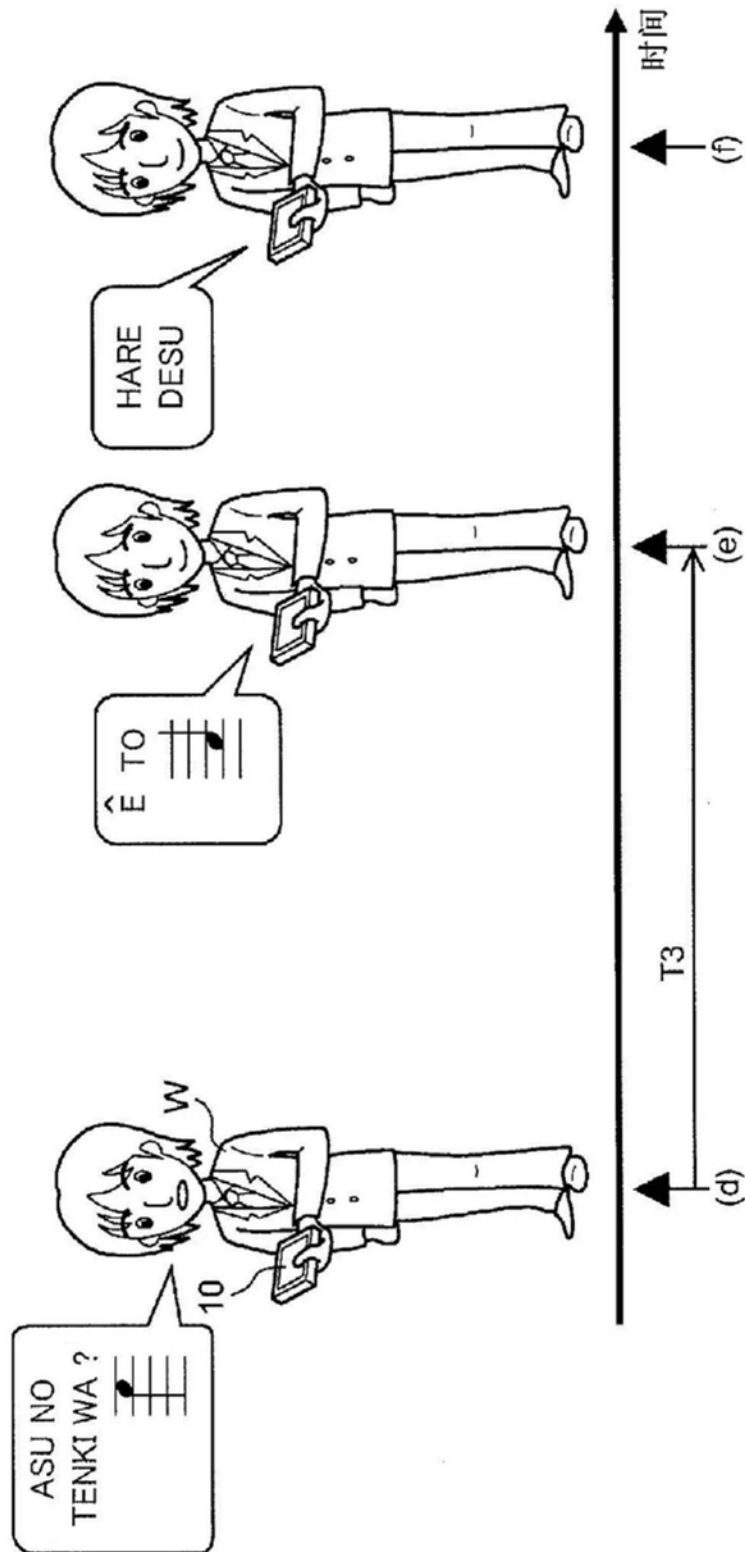


图26

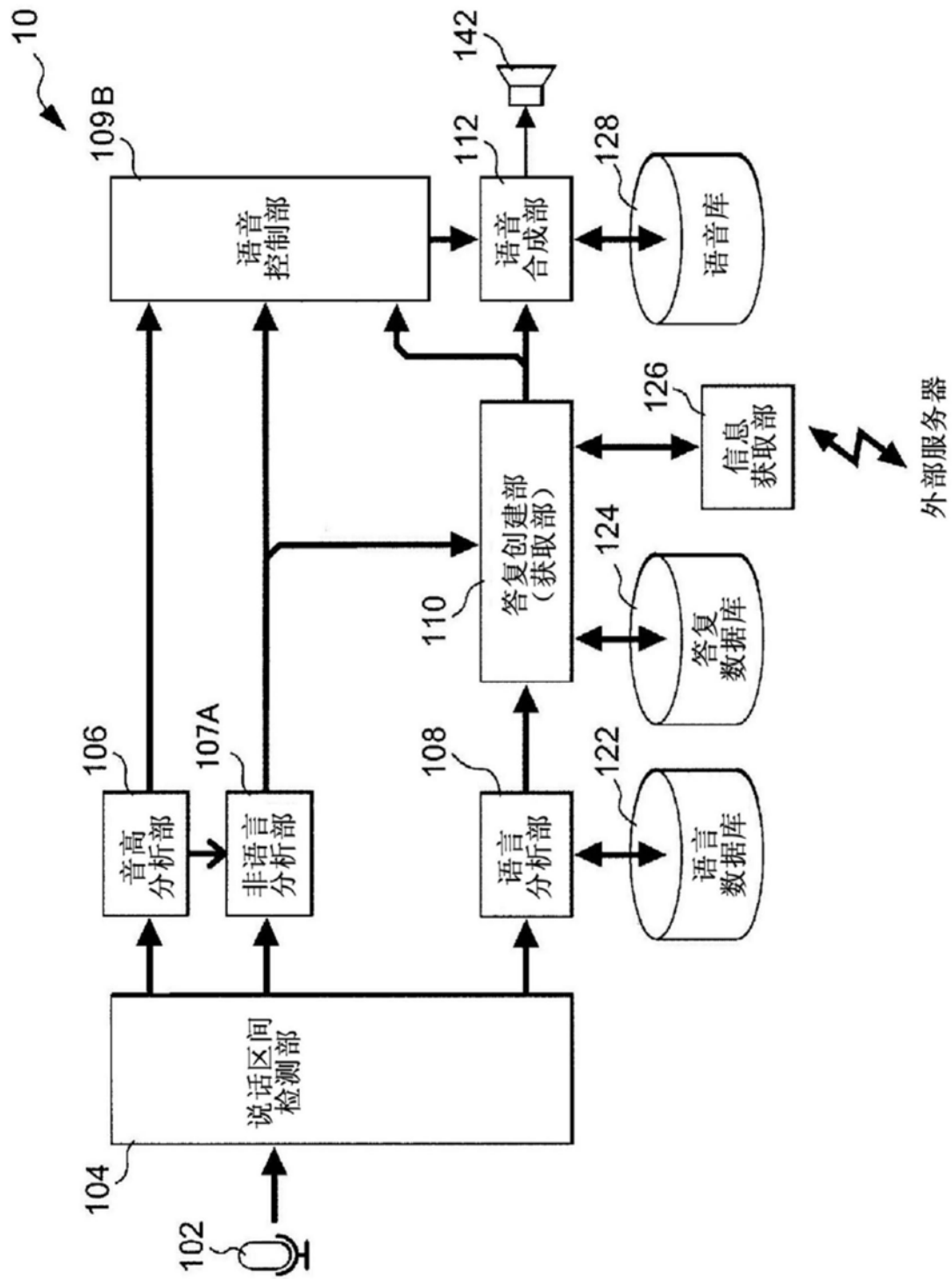


图27

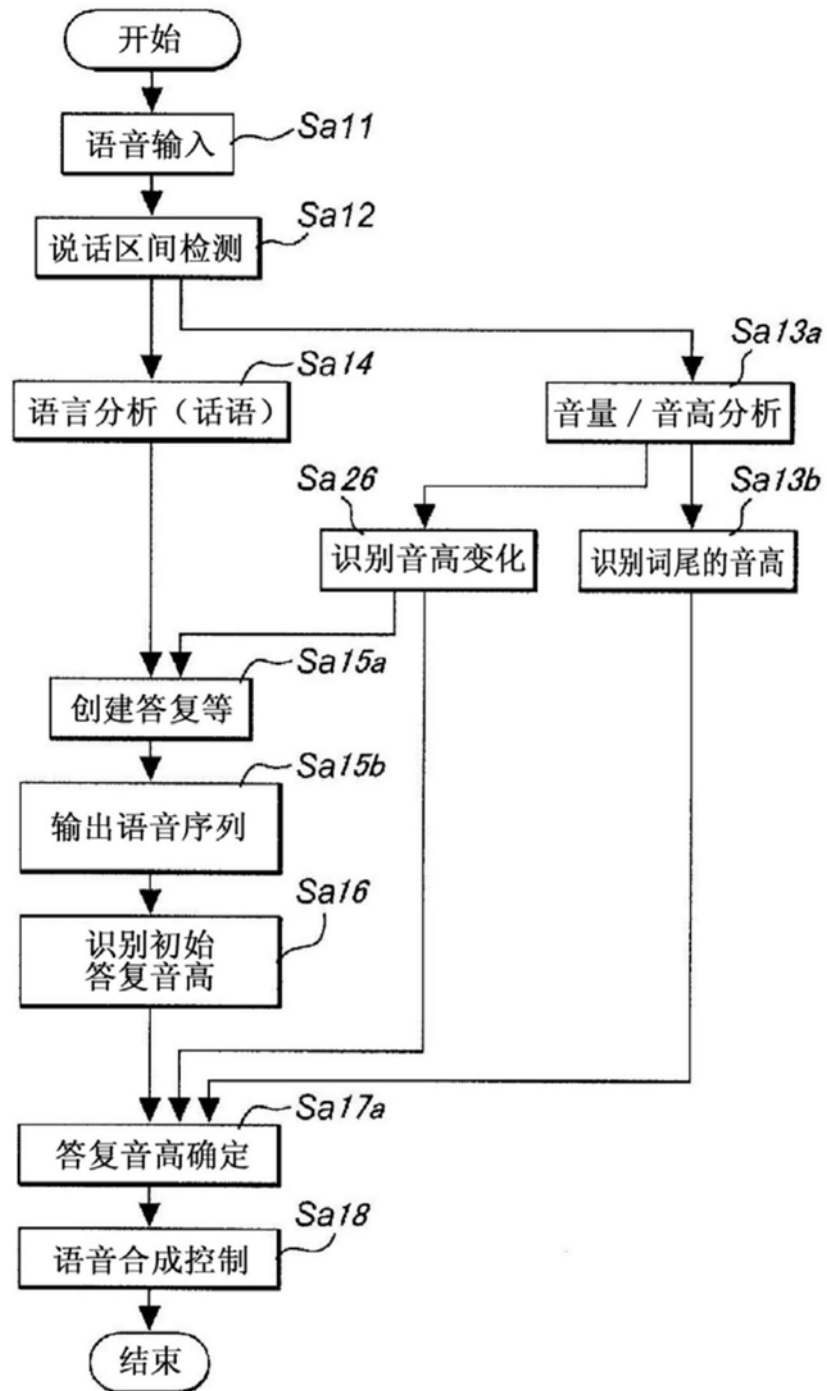


图28

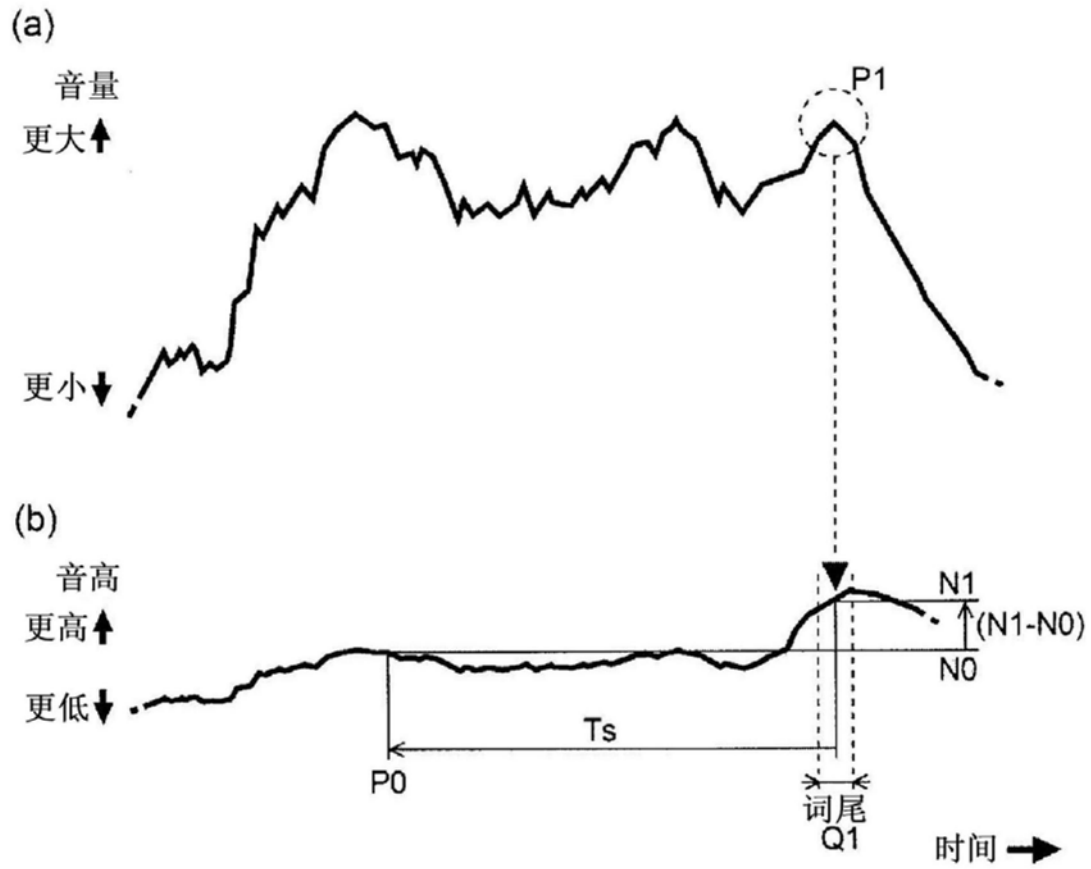


图29

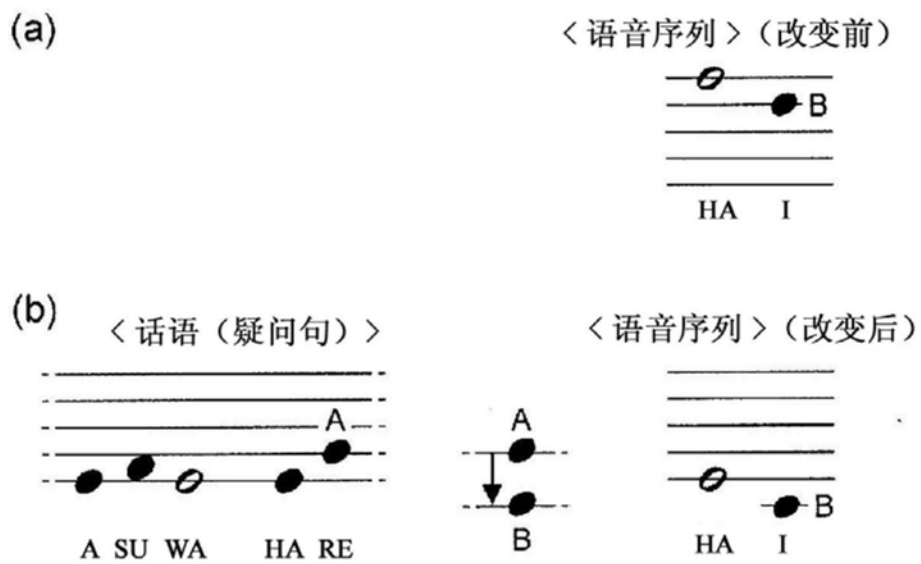
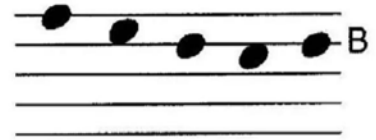


图30

(a)

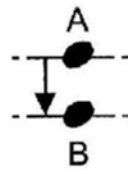
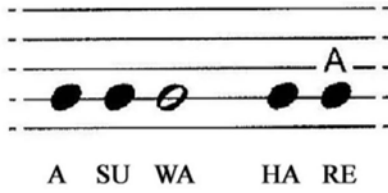
< 语音序列 > (改变前)



SO U DE SU NE

(b)

< 话语 (独白) >



< 语音序列 > (改变后)

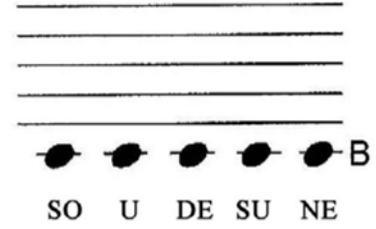


图31

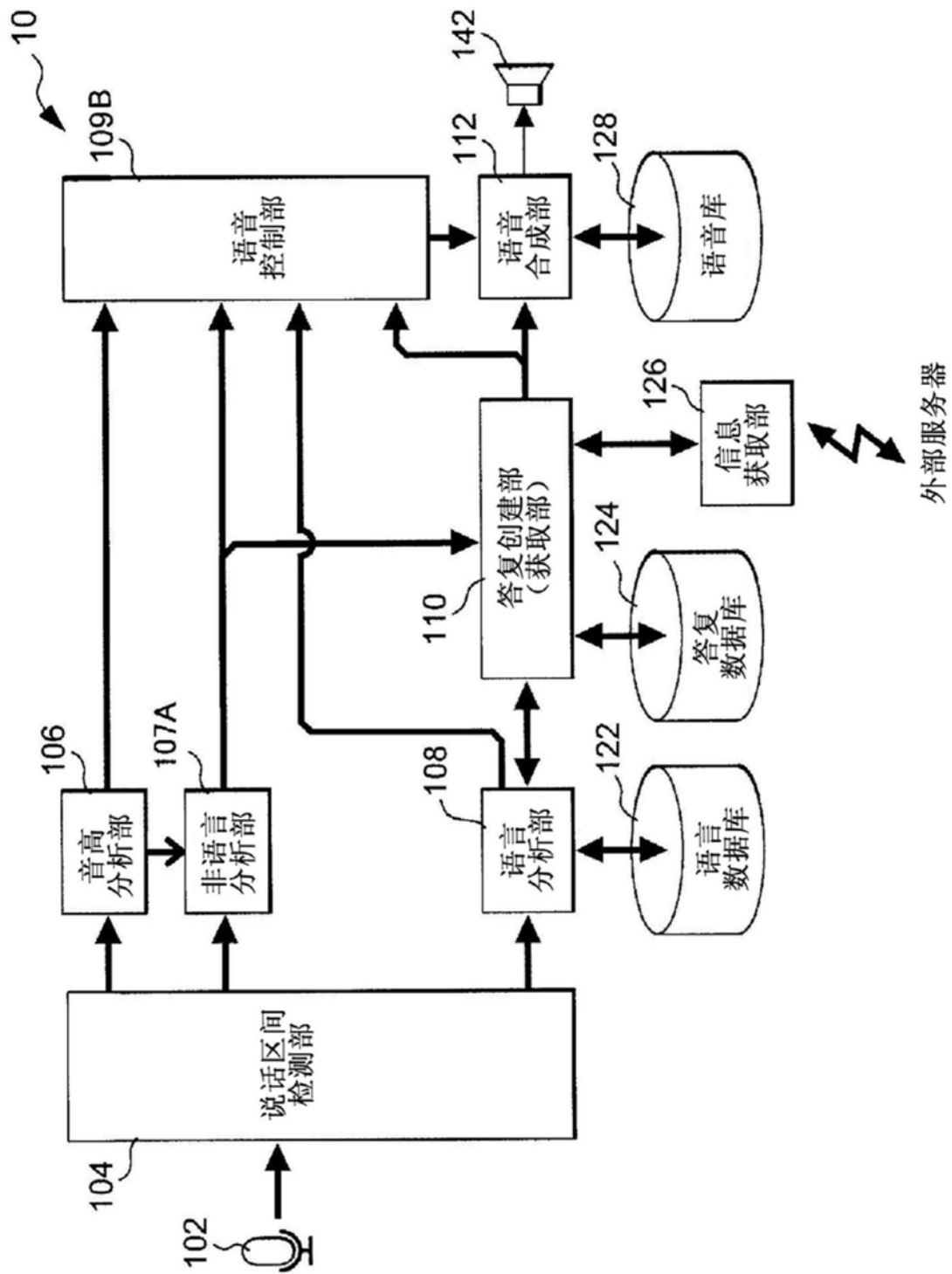


图32

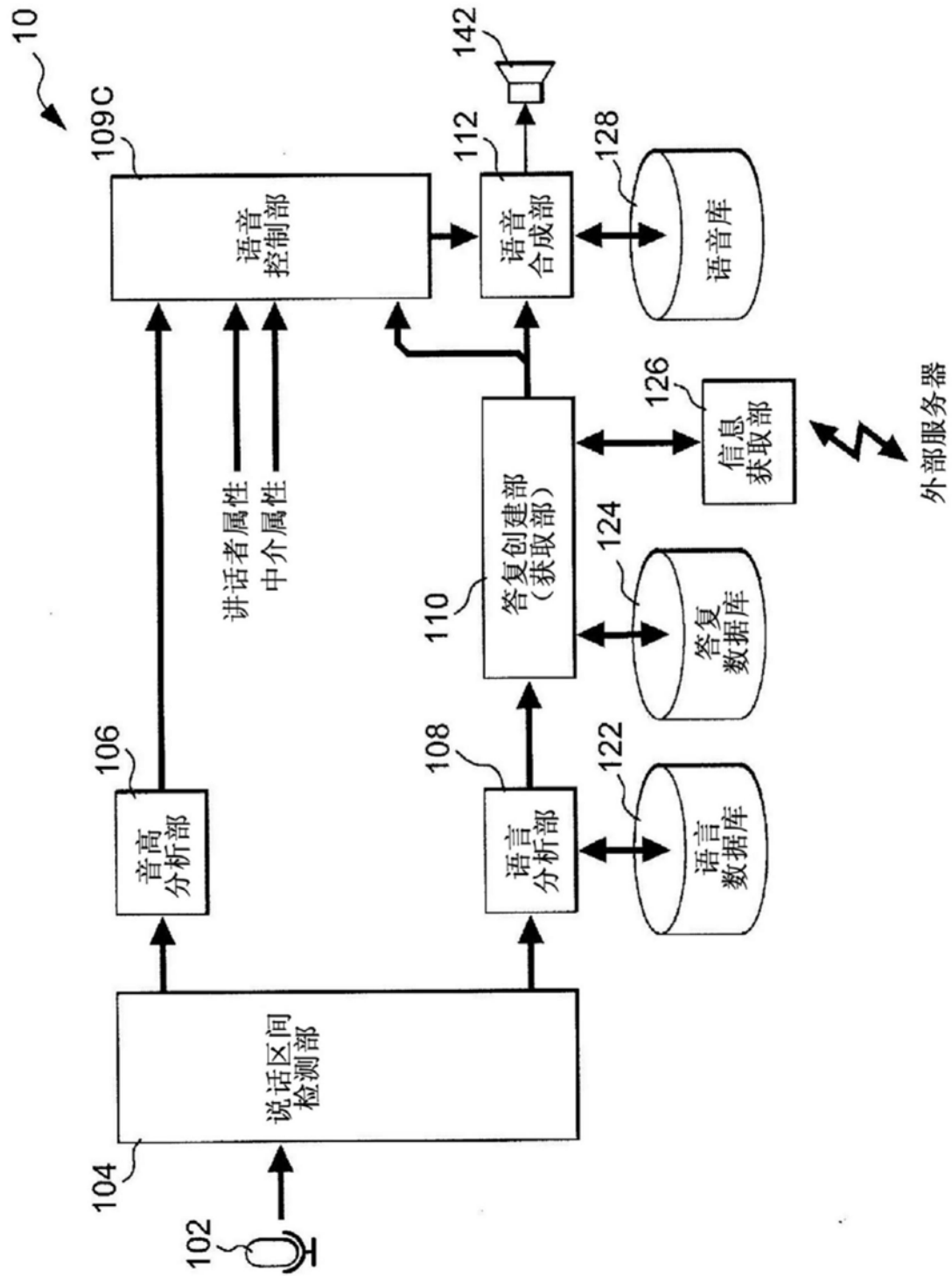


图33

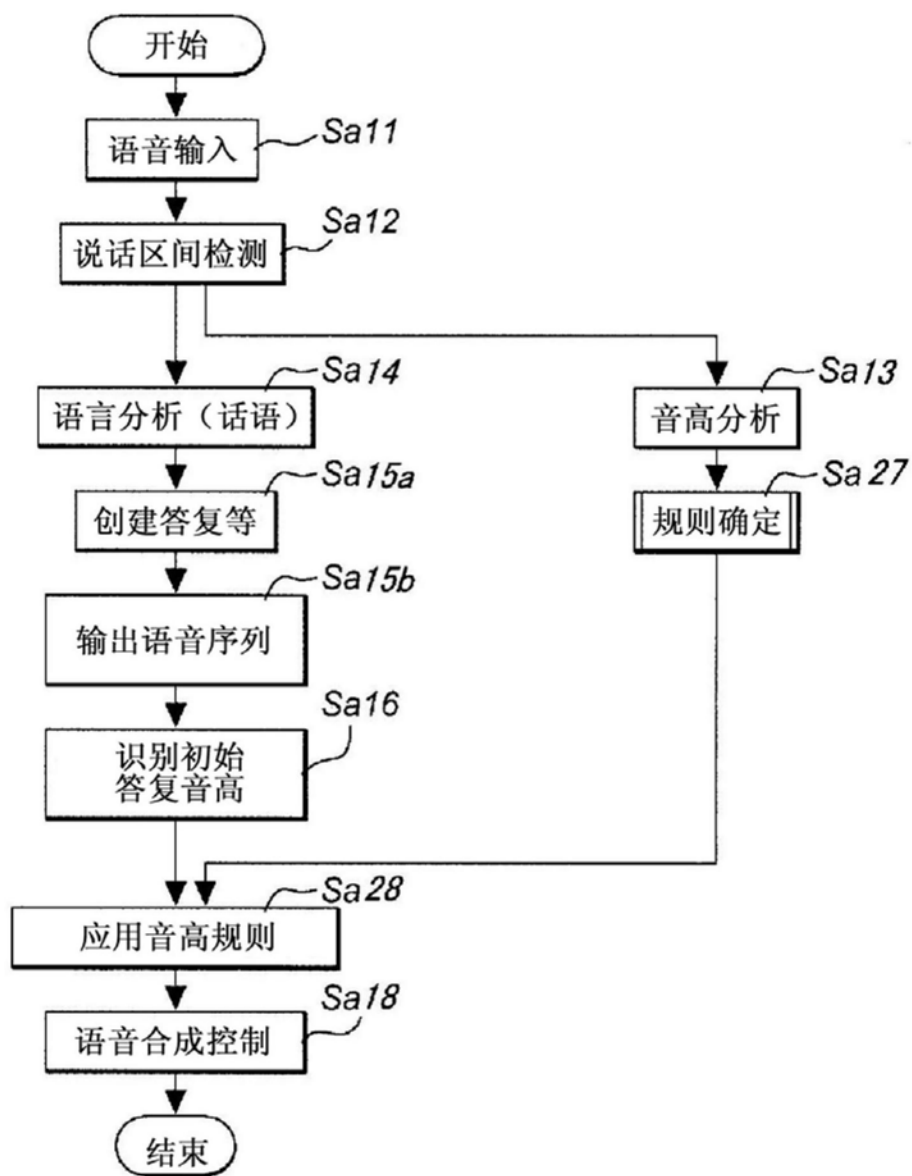


图34

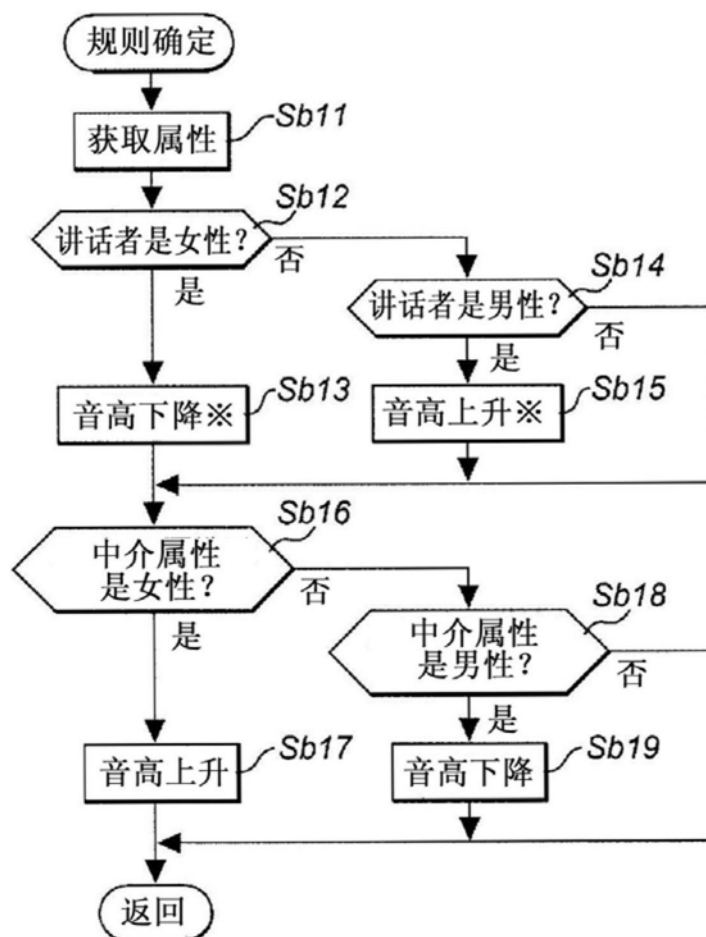
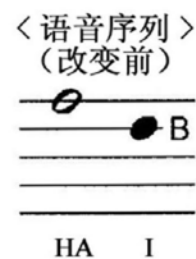


图35

(a)



(b) 〈话语〉

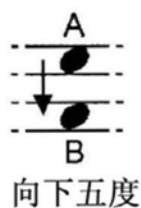


图36

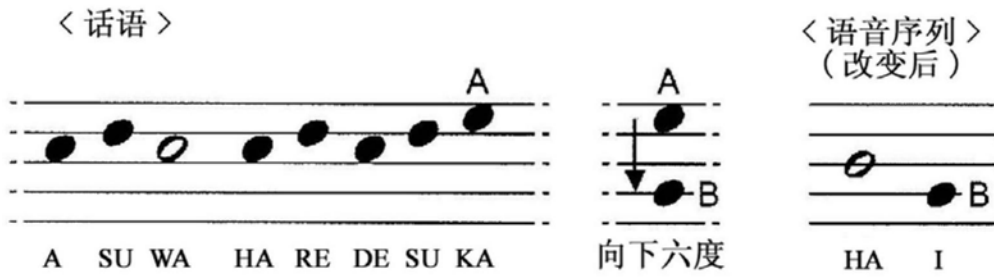


图37

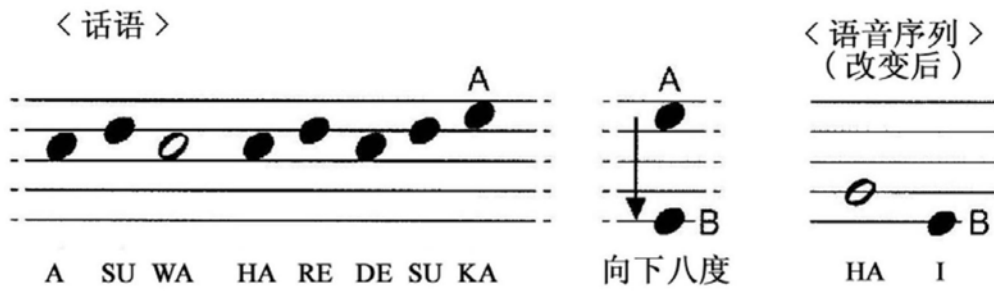


图38

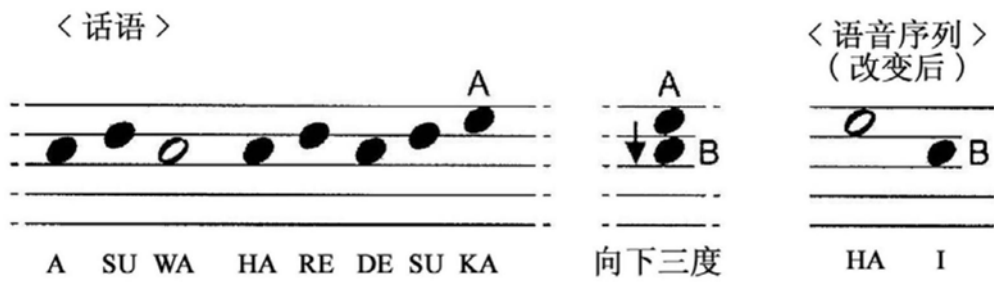


图39



图40

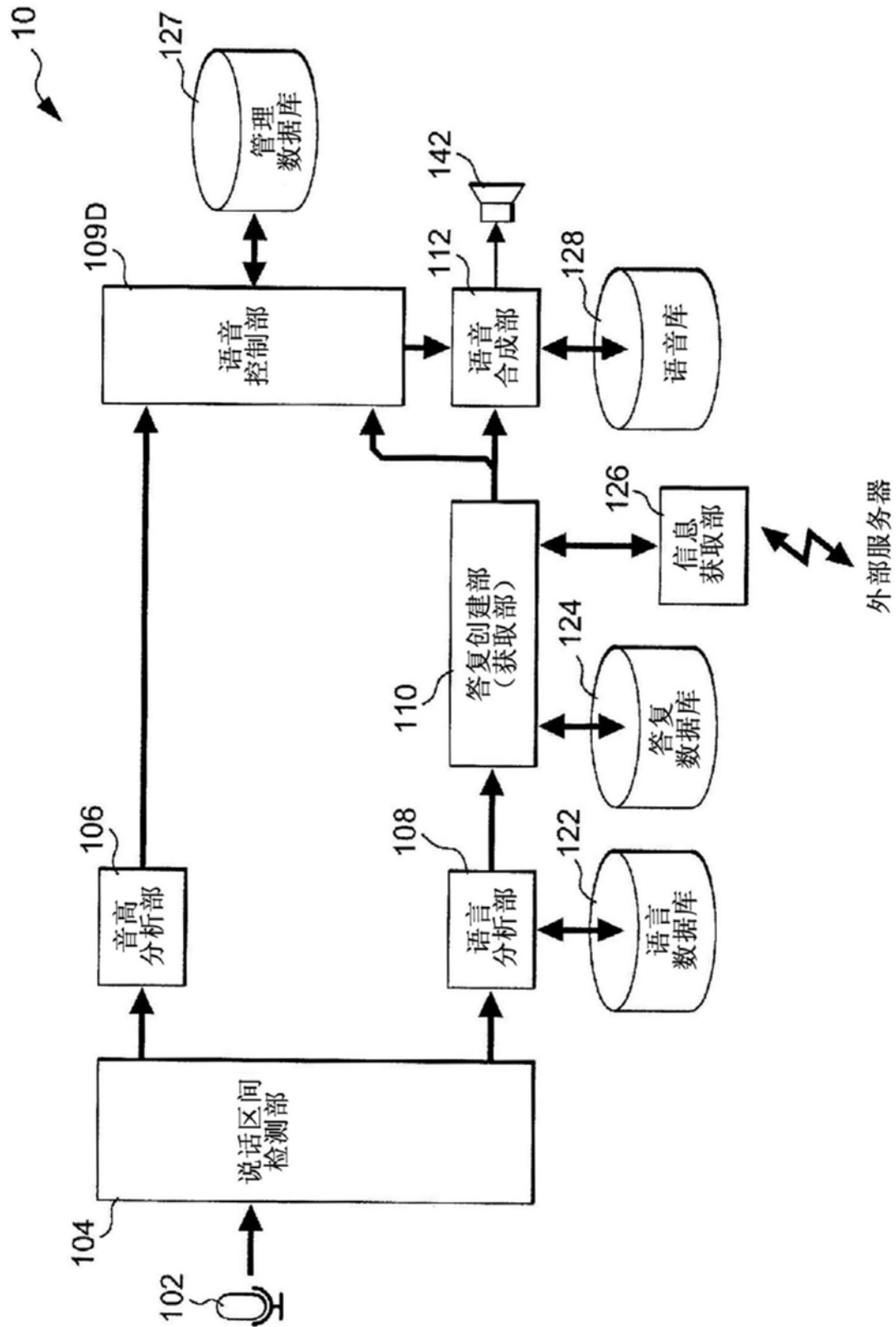


图41

< 指标表（音高规则） >

	音高规则	话语的数量	应用的数量
	向上四度	...	...
▶	向下三度	1 5	1 8
▷	向下五度	1 2	1 8
	向下六度	8	1 7
	向下八度	...	...

图42

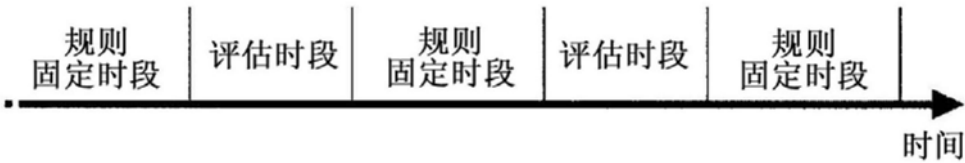


图43

< 语音合成处理 >

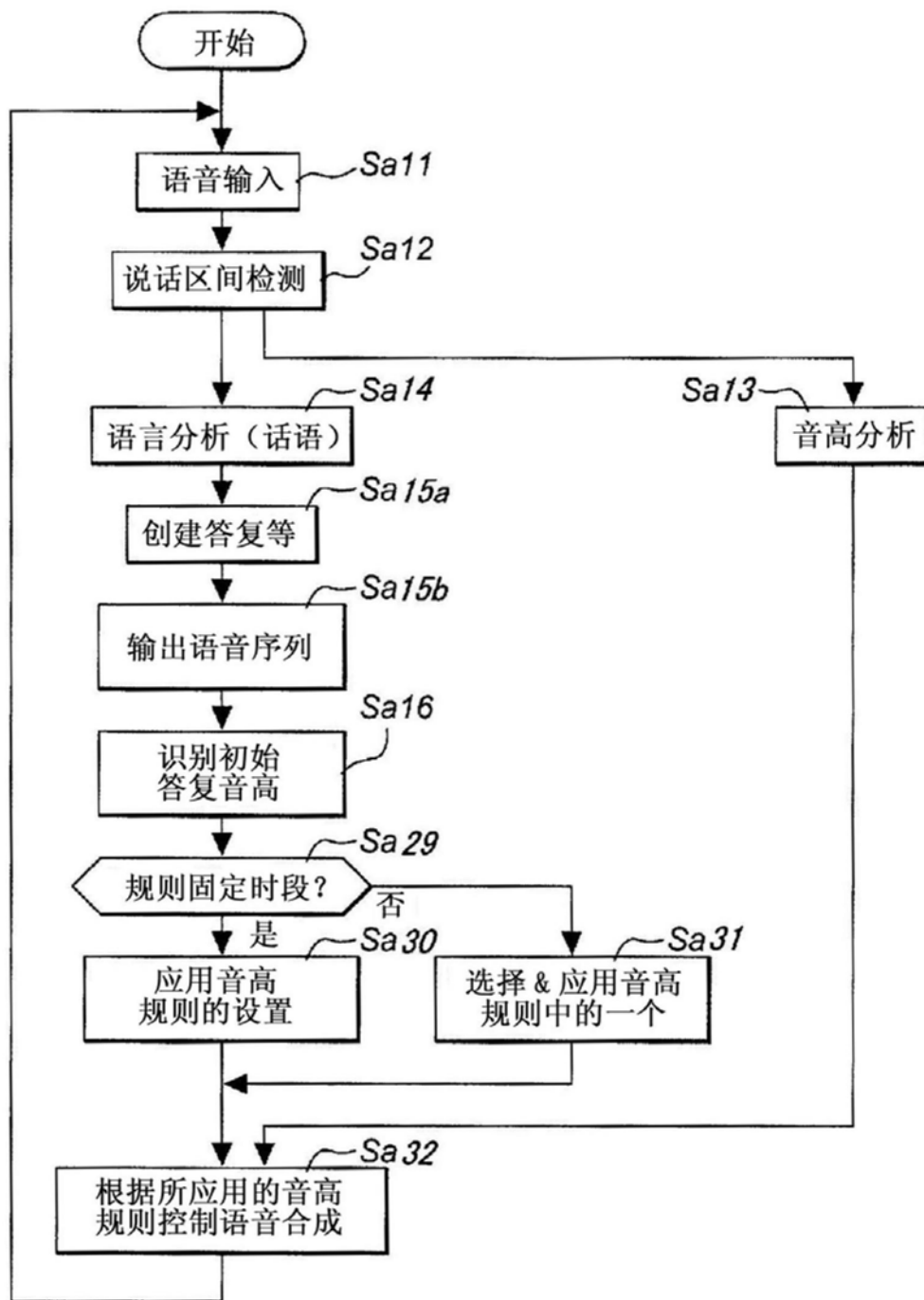


图44

< 表更新处理 >

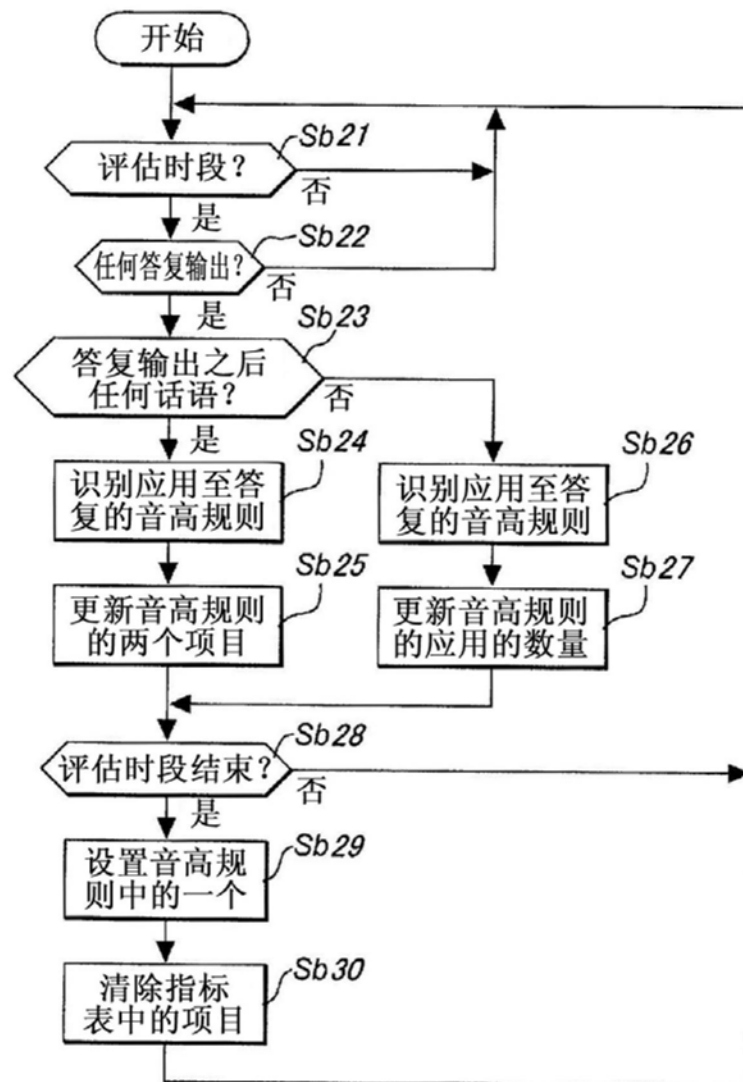


图45

< 指标表 (输出规则) >

输出规则	话语的数量	应用的数量
0. 5秒	...	...
1. 0秒	16	18
1. 5秒	10	18
2. 0秒	9	17
2. 5秒	...	...

图46

< 指标表 (音高规则) >

(a) 讲话者 (说话者): 女性、答复 (设备): 女性

音高规则	话语的数量	应用的数量
向上四度	...	...
向下三度	15	18
▷ 向下五度	12	18
向下六度	8	17
向下八度	...	...

(b) 讲话者: 女性 答复: 男性

音高规则	话语的数量	应用的数量
向上四度	...	...
向下三度	...	...
向下五度	13	18
▷ 向下六度	15	18
向下八度	10	18

(c) 讲话者: 男性 答复: 女性

音高规则	话语的数量	应用的数量
向上四度	...	...
⋮	⋮	⋮
向下八度	...	...

(d) 讲话者: 男性 答复: 男性

音高规则	话语的数量	应用的数量
向上四度	...	...
⋮	⋮	⋮
向下八度	...	...

图47

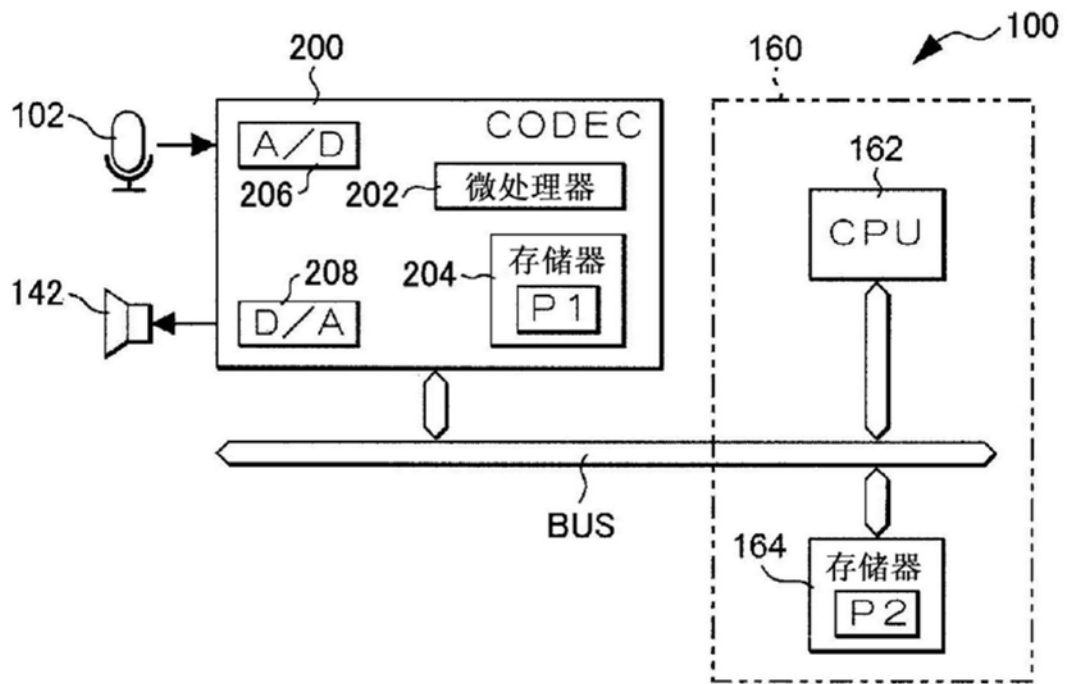


图48

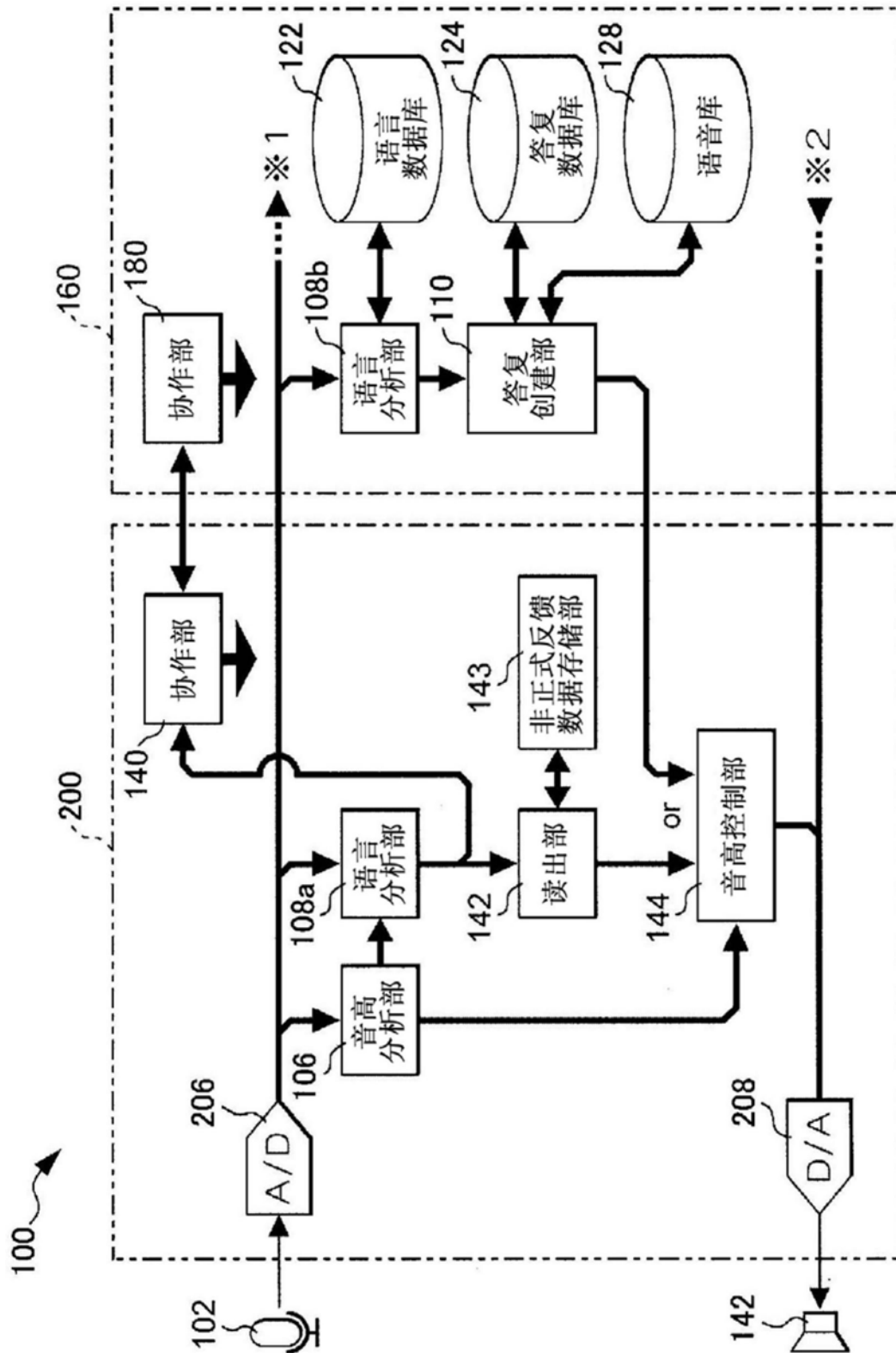


图49

< 语音处理 >

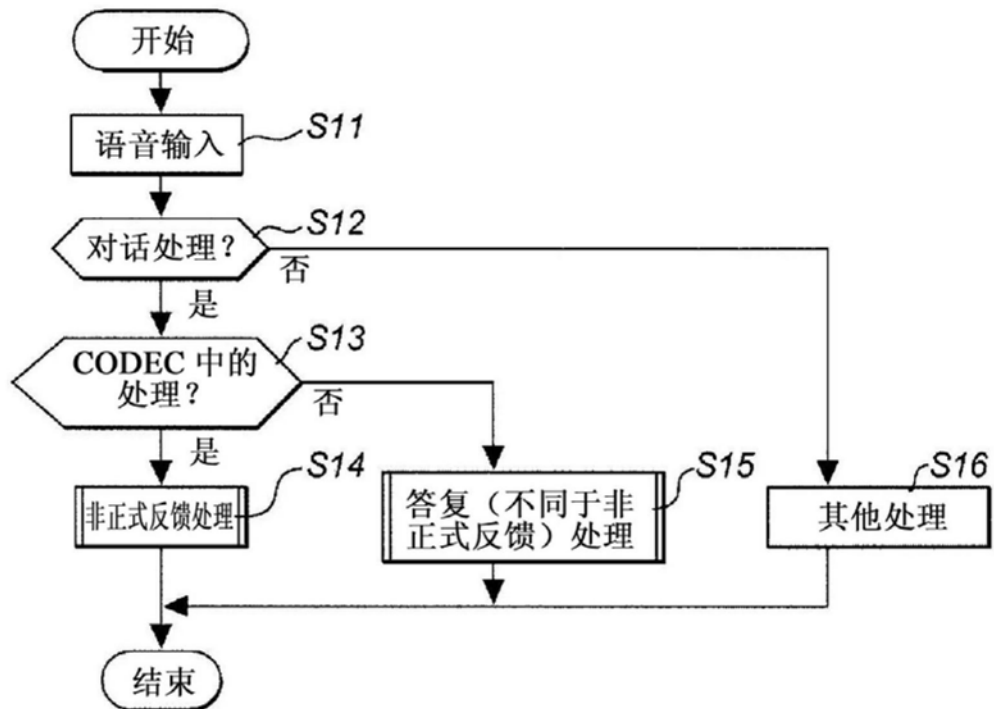


图50

< 非正式反馈处理 > (S14)

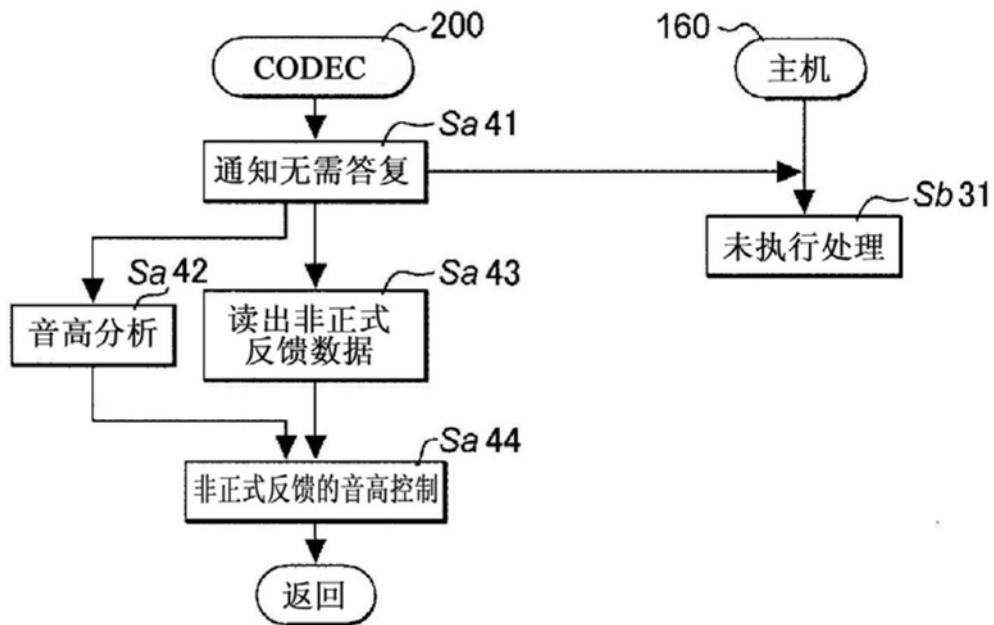


图51

〈答复（不同于非正式反馈）处理〉(S15)

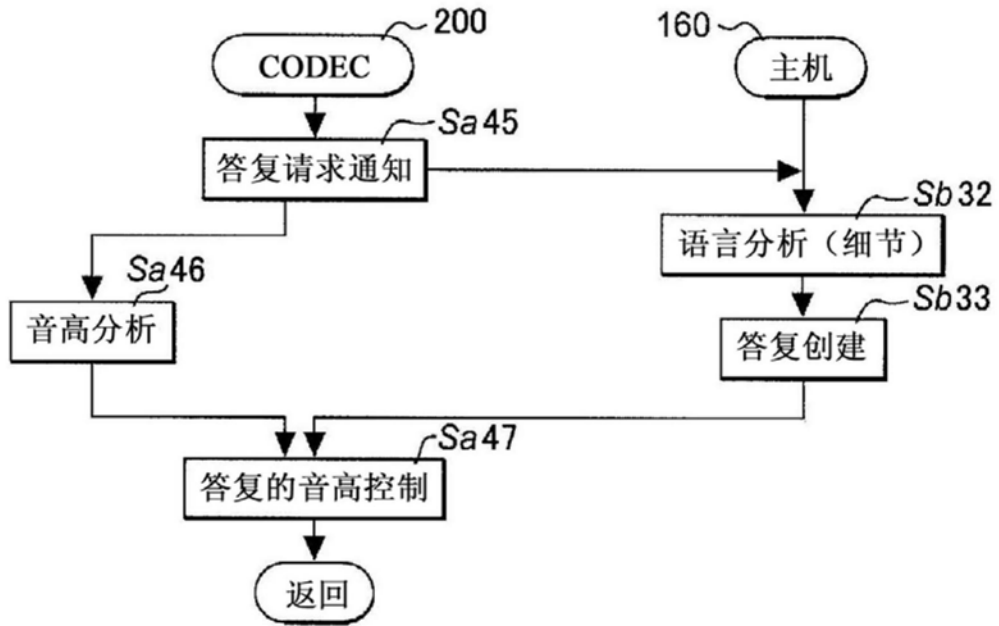


图52