



(12) 发明专利申请

(10) 申请公布号 CN 104488027 A

(43) 申请公布日 2015. 04. 01

(21) 申请号 201380036292. 8

(74) 专利代理机构 中科专利商标代理有限责任公司 11021

(22) 申请日 2013. 07. 01

代理人 李逸雪

(30) 优先权数据

2012-153941 2012. 07. 09 JP

(51) Int. Cl.

G10L 25/69(2013. 01)

(85) PCT国际申请进入国家阶段日

2015. 01. 07

(86) PCT国际申请的申请数据

PCT/JP2013/068007 2013. 07. 01

(87) PCT国际申请的公布数据

W02014/010450 JA 2014. 01. 16

(71) 申请人 独立行政法人情报通信研究机构

地址 日本东京都

(72) 发明人 杉浦孔明 大熊英男 木村法幸

志贺芳则 林辉昭 水上悦雄

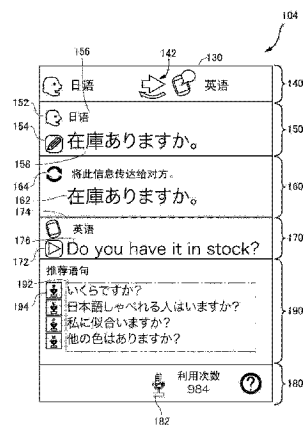
权利要求书2页 说明书15页 附图8页

(54) 发明名称

声音处理系统以及终端装置

(57) 摘要

本发明提供一种能够提高声音识别的精度且用户容易利用的声音处理系统。声音处理系统如果接受了讲话声音,则进行声音识别,并显示识别结果的文本(158)。声音处理系统进一步按照设定来将该识别结果翻译成其他语言的文本(176)后进行显示,并且产生其合成声音。声音处理系统进一步使用讲话时的各种传感器的输出、预先完成学习的讲话连续模型、讲话候补的翻译以及声音识别得分,选择作为下一次讲话讲出的概率高且翻译以及声音识别得分也高的讲话候补,作为讲话候补推荐列表(192)来进行推荐。用户能够以该讲话候补推荐列表(192)内的讲话作为启发,考虑下一次讲话。



1. 一种声音处理系统,包括:

讲话输入部件,用于接受讲话信息的输入,该讲话信息包含表示讲话的声音信号和表示完成该讲话的环境的规定环境信息;

声音识别部件,用于对所述讲话输入部件接受的讲话信息内的声音信号进行声音识别,并将识别结果作为文本来输出;

数据处理部件,对所述声音识别部件输出的文本执行规定的数据处理;

讲话连续模型存储部件,用于存储以统计方式学习完毕的讲话连续模型,统计方式学习完毕的讲话连续模型如下:如果接受了讲话的文本和所述规定环境信息,则能够针对规定的讲话集合之中的讲话,计算出由所述文本表示的讲话之后连续讲出的概率;

讲话存储部件,用于存储所述数据处理部件对所述规定的讲话集合内的讲话和针对该讲话集合内的各个讲话进行了处理时的数据处理的可靠度;以及

讲话候补推荐部件,使用所述声音识别部件对所述讲话输入部件接受的讲话信息的识别结果和该讲话信息中包含的环境信息,基于以规定的形式将使用存储在所述讲话连续模型存储部件中的所述讲话连续模型对所述规定集合内的各讲话计算出的概率、和存储在所述讲话存储部件中的针对所述规定集合内的各讲话的所述数据处理的可靠度进行组合而得到的评价得分,在所述多个讲话的集合内,对推荐给进行了所述声音识别部件识别出的讲话的用户的讲话候补附加得分,并基于该得分而向用户推荐讲话候补。

2. 根据权利要求 1 所述的声音处理系统,其中,

所述数据处理部件包括自动翻译部件,该自动翻译部件受理接受某讲话后所述声音识别部件输出的识别结果,将该识别结果自动翻译成与所述某讲话的语言不同的其他语言,并作为文本来输出,

所述可靠度是该自动翻译部件的翻译结果作为相对于所述某讲话的所述其他语言的翻译的似然性。

3. 根据权利要求 2 所述的声音处理系统,其中,

所述数据处理部件还包括声音合成部件,该声音合成部件用于基于所述自动翻译部件输出的所述其他语言的文本来对所述其他语言的声音信号进行合成。

4. 根据权利要求 1~3 中任一项所述的声音处理系统,其中,

所述讲话候补推荐部件包括以下部件,该部件用于基于使用所述讲话连续模型对所述规定集合内的各讲话计算出的概率、与存储在所述讲话存储部件中的针对所述规定集合内的各讲话的可靠度之间的线性和的评价,在所述多个讲话的集合内推测所述声音识别部件进行声音识别的讲话之后的后续讲话的候补,

所述线性和中的所述可靠度和所述概率的系数都是正数。

5. 根据权利要求 1~4 中任一项所述的声音处理系统,其中,

所述声音处理系统还包括讲话候补提示部件,该讲话候补提示部件用于向用户提示由所述讲话候补推荐部件推荐的讲话候补。

6. 根据权利要求 1~5 中任一项所述的声音处理系统,其中,

所述声音处理系统还包括讲话文本信息输入部件,该讲话文本信息输入部件用于接受包含表示讲话的文本和所述规定环境信息在内的讲话文本信息的输入,取代所述声音识别部件的输出而向所述讲话候补推荐部件以及所述数据处理部件提供该讲话文本信息内的

文本。

7. 一种终端装置,包括:麦克风;用于收集与周围环境相关的信息的传感器的集合;显示装置;通信装置;以及讲话信息发送部件,与所述麦克风、所述传感器的集合以及所述通信装置连接,用于经由所述通信装置,将讲话信息发送到规定的声音处理服务器,并委托声音识别和针对识别结果的规定的数据处理,其中,所述讲话信息包含从所述麦克风接受讲话后输出的信号中得到的声音信号和从得到该声音信号时的所述传感器的集合得到的信息,

所述终端装置还包括:

处理结果提示部件,与所述通信装置连接,且用于接受响应所述委托而从所述声音处理服务器发送来的所述数据处理的处理结果,并向用户提示该处理结果;以及

讲话候补推荐列表显示部件,从所述声音处理服务器接受被推荐为多个讲话候补的讲话候补推荐列表,并通过显示于所述显示装置来向用户推荐讲话候补。

8. 根据权利要求 7 所述的终端装置,其中,

所述声音处理服务器对所述声音识别的结果进行的所述规定的数据处理是以下处理:将所述讲话自动翻译成与该讲话的语言不同的其他语言,并进一步对该自动翻译的结果的声音进行合成,

从所述声音处理服务器发送的所述数据处理的处理结果是表示由所述声音处理服务器合成的声音的声音信号,

所述处理结果提示部件包括:扬声器;以及用于根据表示由所述声音处理服务器合成的声音的声音信号来驱动所述扬声器的部件。

9. 根据权利要求 7 或 8 所述的终端装置,其中,

所述终端装置还包括:

选择部件,用户能够为了选择由所述讲话候补推荐列表显示部件显示的讲话候补当中的任一个讲话候补,而操作该选择部件;以及

讲话文本信息发送部件,响应于由所述选择部件选择出所述讲话候补推荐列表内的讲话候补当中的任一个讲话候补的情况,经由所述通信装置,将包含选择出的讲话候补的文本和从所述传感器的集合得到的信息在内的讲话文本信息发送到规定的声音处理服务器,并委托针对该讲话文本信息进行所述规定的数据处理。

声音处理系统以及终端装置

技术领域

[0001] 本发明涉及使用了声音识别的服务,特别涉及户能够顺畅地进行使用了声音识别的交流的的技术。

背景技术

[0002] 伴随着移动电话特别是所谓智能电话的普及,出现了各种应用程序。其中,在输入中使用声音的应用程序被认为在今后将进一步得到普及。这是因为在智能手机这样小的装置中存在难以进行文本输入问题。

[0003] 在输入中使用声音的应用程序之中,若是仅朗读“是 / 否 (はい / いいえ)”等显示于画面的字符串的应用程序,用户不会对应当说什么而感到迷茫。但是,在例如如声音翻译这样将声音输入用作与他人进行交流的工具的情况下,由于想不到应该用声音输入什么样的内容才好,所以用户有时会对应用程序的使用感到犹豫。例如,考虑用户尝试使用声音翻译的应用程序的情况。在做尝试时,经常会经历到用户没有认真思考应当说的语言。

[0004] 进一步地,大多情况下在实际发出声音之前,会发出“嗯”、“啊”之类的没有含义的声音。在声音识别中,这样没有含义的声音会降低声音识别的精度。

[0005] 如上所述,虽然要说些什么但大多情况下却想不到说些什么才好这样的问题并不限于声音翻译。在多人会话中也可能存在同样的情况。例如,存在以下情况:虽然正在进行会话,但是话题中断,想不到接下来说些什么才好。

[0006] 在专利文献 1 中公开了一种话题提供装置,用于在多人的会话中,当会话的话题中断时提供话题。该专利文献 1 公开的系统使用话题数据库,该话题数据库预先将话题和与该话题相关的关键字建立关联后存储,其中,关键字来自于被共同用于多个话题中的观点。准备多个观点。该系统从用户的会话中提取词句,针对多个观点中的每一个观点,计算其权重,进一步地,针对在最后提示给用户的话题和存储在话题数据库中的话题,按每个观点计算出相似度,将该相似度乘以针对各观点计算出的权重,并计算出这些乘法值之和。然后,将该和最大的话题作为新话题而提示给用户。作为多个观点,可列举相当于所谓“5W1H”的观点。

[0007] 根据专利文献 1,通过该装置,能够提供与成为当前会话的中心的观点相接近而对于其他观点来说相差较多的话题。

[0008] 在先技术文献

[0009] 专利文献

[0010] 专利文献 1 :JP 特开 2011-123530 号公报

发明内容

[0011] 发明要解决的课题

[0012] 但是,在专利文献 1 记载的发明中,虽然如果会话的内容丰富到一定程度就能够确定话题,但是如果会话较短,则存在难以决定应当提供的话题的问题。特别是,在使用了

声音翻译的会话这样讲话非常短的情况下,很难应用专利文献 1 记载的发明。

[0013] 进一步,作为在声音翻译中特有的问题,存在讲话人过多地考虑如何讲话才能正确地进行声音识别并正确地进行翻译的问题。为此,会迷惑如何讲、讲些什么,其结果,大多情况下都会降低声音翻译的精度。当然,如果用户说了容易声音识别且容易自动翻译的话,则对于声音翻译来说是优选的。但是,为此,会使讲话本身变得困难,所以不能充分地发挥声音翻译的效果。

[0014] 因此,本发明提供一种能够提高声音识别的精度且用户容易利用的声音处理系统以及由这样的声音处理系统利用的终端装置。

[0015] 本发明的其他目的是提供一种能够提高声音识别的精度和利用声音识别的结果来提供的声音服务的精度、且用户容易利用的声音处理系统、以及由这样的声音处理系统利用的终端装置。

[0016] 用于解决课题的手段

[0017] 本发明的第 1 方面涉及的声音处理系统包括:讲话输入部件,用于接受讲话信息的输入,该讲话信息包含表示讲话的声音信号和表示完成该讲话的环境的规定环境信息;声音识别部件,用于对讲话输入部件接受的讲话信息内的声音信号进行声音识别,并将识别结果作为文本来输出;数据处理部件,对声音识别部件输出的文本执行规定的数据处理;讲话连续模型存储部件,用于存储以统计方式学习完毕的讲话连续模型,统计方式学习如下:如果接受了讲话的文本和规定环境信息,则能够针对规定的讲话集合之中的讲话,计算出由文本表示的讲话之后连续讲出的概率;讲话存储部件,用于存储数据处理部件对规定的讲话集合内的讲话和针对该讲话集合内的各个讲话进行了处理时的数据处理的可靠度;以及讲话候补推荐部件,使用讲话输入部件对声音识别部件接受的讲话信息的识别结果和该讲话信息中包含的环境信息,基于以规定的形式将使用存储在讲话连续模型存储部件中的讲话连续模型对规定集合内的各讲话计算出的概率、和存储在讲话存储部件中的针对规定集合内的各讲话的数据处理的可靠度进行组合而得到的评价得分,在多个讲话的集合内,对推荐给进行了声音识别部件识别出的讲话的用户的讲话候补附加得分,并基于该得分而向用户推荐讲话候补。

[0018] 如果讲话输入部件接受讲话信息的输入,则声音识别部件进行针对讲话信息中的声音信号的声音识别,并输出识别结果的文本。数据处理部件对该文本进行数据处理。使用识别结果的文本和包含在讲话信息中的环境信息,讲话候补推荐部件借助讲话连续模型来计算出各讲话成为下一次讲话的概率,进一步使用该概率和数据处理中的各讲话的可靠度,进行推荐为下一次讲话的讲话的得分附加,根据该结果推荐下一次讲话候补。

[0019] 由于推荐了下一次讲话的候补,所以用户能够将该讲话候补作为启发来考虑下一次讲话。由此,能够降低想不出该讲什么而使得用户感到不知所措的事态发生的危险性。

[0020] 优选,数据处理部件包括自动翻译部件,该自动翻译部件受理接受某讲话后声音识别部件输出的识别结果,将该识别结果自动翻译成与某讲话的语言不同的其他语言,并作为文本来输出。可靠度是该自动翻译部件的翻译结果作为相对于某讲话的其他语言的翻译的似然性。

[0021] 更优选,数据处理部件还包括声音合成部件,该声音合成部件用于基于自动翻译部件输出的其他语言的文本来对其他语言的声音信号进行合成。

[0022] 讲话候补推荐部件也可以包括以下部件,该部件用于基于使用讲话连续模型对规定集合内的各讲话计算出的概率、与存储在讲话存储部件中的针对规定集合内的各讲话的可靠度之间的线性和的评价,在多个讲话的集合内推测声音识别部件进行声音识别的讲话之后的后续讲话的候补。此时,线性和中的可靠度和概率的系数都是正数。

[0023] 优选,声音处理系统还包括讲话候补提示部件,该讲话候补提示部件用于向用户提示由讲话候补推荐部件推荐的下一轮的讲话候补。

[0024] 更优选,声音处理系统还包括讲话文本信息输入部件,该讲话文本信息输入部件用于接受包含表示讲话的文本和规定环境信息在内的讲话文本信息的输入,取代声音识别部件的输出而将该讲话文本信息内的文本提供至讲话候补推荐部件以及数据处理部件。

[0025] 本发明的第2方面涉及的终端装置包括:包括:麦克风;用于收集与周围环境相关的信息的传感器的集合;显示装置;通信装置;以及讲话信息发送部件,与麦克风、传感器的集合以及通信装置连接,用于经由通信装置,将讲话信息发送到规定的声音处理服务器,并委托声音识别和针对识别结果的规定的数据处理,其中,讲话信息包含从麦克风接受讲话后输出的信号中得到的声音信号和从得到该声音信号时的传感器的集合得到的信息。终端装置还包括:处理结果提示部件,与通信装置连接,且用于接受响应委托而从声音处理服务器发送来的数据处理的处理结果,并向用户提示该处理结果;以及讲话候补推荐列表显示部件,从声音处理服务器接受被推荐为多个讲话候补的讲话候补推荐列表,并通过显示于显示装置来向用户推荐讲话候补。

[0026] 优选,声音处理服务器对声音识别的结果进行的规定的数据处理是以下处理:将讲话自动翻译成与该讲话的语言不同的其他语言,并进一步对该自动翻译的结果的声音进行合成。从声音处理服务器发送的数据处理的处理结果是表示由声音处理服务器合成的声音的声音信号。处理结果提示部件包括:扬声器;以及用于根据表示由声音处理服务器合成的声音的声音信号来驱动扬声器的部件。

[0027] 进一步优选,终端装置还包括:选择部件,用户能够为了选择由讲话候补推荐列表显示部件显示的讲话候补当中的任一个讲话候补,而操作该选择部件;以及讲话文本信息发送部件,响应于由选择部件选择出讲话候补推荐列表内的讲话候补当中的任一个讲话候补的情况,经由通信装置,将包含选择出的讲话候补的文本和从传感器的集合得到的信息在内的讲话文本信息发送到规定的声音处理服务器,并委托针对该讲话文本信息进行规定的数据处理。

[0028] 发明效果

[0029] 如以上所述,根据本发明,能够减小利用声音服务的用户对讲些什么没有想法而感到不知所措的可能性,并能够使声音服务更易于使用。并且,在想要讲话之前发出的无意义的声音变少,能够提高声音识别的精度,也能够提高使用该声音识别结果的声音服务的精度。

[0030] 其结果,能够提供一种能提高声音识别的精度并且用户易于利用的声音处理装置以及终端装置。

[0031] 进一步地,能够提供一种能提高声音识别的精度和使用声音识别结果的声音服务的精度并且用户易于利用的声音处理装置以及终端装置。

附图说明

[0032] 图 1 是示意性表示本发明的第 1 实施方式涉及的声音翻译系统的整体结构的图。

[0033] 图 2 是示意性表示显示于在图 1 所示的系统中所使用的便携式终端的画面上的用于声音翻译的画面的图。

[0034] 图 3 是表示在第 1 实施方式的声音翻译系统中在便携式终端与服务器之间进行的用于声音翻译的处理序列的图。

[0035] 图 4 是表示便携式终端的硬件结构的框图。

[0036] 图 5 是表示便携式终端中实现伴随使用了声音输入的声音翻译的处理的程序的控制构造的流程图。

[0037] 图 6 是在第 1 实施方式的声音翻译系统中执行接受来自多个便携式终端的声音输入并翻译成指定的语言,进一步对其声音进行合成后送回便携式终端的处理的服务器的功能性框图。

[0038] 图 7 是用于实现图 6 所示的服务器的功能的程序的流程图。

[0039] 图 8 是在图 6 所示的服务器中使用的进行讲话连续模型的学习和文集内的讲话的翻译得分的计算的模型生成部的框图。

具体实施方式

[0040] 在以下的说明以及附图中,对同一部件赋予相同的参照符号。因此,不重复对这些部件的详细说明。

[0041] [第 1 实施方式]

[0042] < 结构 >

[0043] - 整体结构 -

[0044] 参照图 1,本发明涉及的声音翻译系统 100 包括:与因特网 102 连接且进行声音翻译服务的服务器 106;以及能与因特网 102 连接且安装有用于利用声音翻译服务的应用程序的便携式终端 104。

[0045] - 应用程序画面 -

[0046] 参照图 2,便携式终端 104 的用于利用声音翻译服务的应用程序画面 130 被大体分割成 6 个区域。即,用于显示成为声音翻译服务的对象的语言对(源语言和目标语言)的语言显示区域 140;用于显示以源语言的声音输入的语句的声音识别结果或文本输入结果的输入文本显示区域 150;显示对声音识别后的语句进行自动翻译而得到的结果、即文本的翻译结果显示区域 170;显示将翻译结果逆翻成原语言后的语句的逆翻译区域 160;显示推荐为下一次的讲话候补的讲话候补的列表(讲话候补推荐列表)的讲话候补推荐区域 190;以及显示声音翻译系统的利用状况的状态区域 180。

[0047] 在语言显示区域 140,分别采用源语言的文字将源语言的语言名显示在左侧,将目标语言的语言名显示在右侧。另外,在应用程序画面 130 中,翻译结果的语句以外的文本都采用源语言的文字来显示。在源语言名以及目标语言名之间,显示用于设定声音翻译的语言的组合的设定按钮 142。

[0048] 在输入文本显示区域 150 显示:源语言的语言名的显示 156;在进行声音输入时由用户操作的声音输入按钮 152;以及用于显示不是采用声音输入而是直接对输入语句的文

本进行输入的文本输入画面的文本输入按钮 154。声音输入的结果以及文本输入的结果都在输入文本显示区域 150 内被显示为输入文本 158。另外,在本实施方式中,在按压声音输入按钮 152 的期间,对声音进行录制。如果中断声音输入按钮 152 的按压,则结束声音的录制。

[0049] 在逆翻译区域 160 显示:将根据声音输入的结果来生成并被自动翻译出的目标语言的语句逆翻成源语言的语句而得到的结果的语句 162;以及在对语句进行文本输入等时用于开始该语句的翻译的翻译按钮 164。通过将所输入的源语言的语句得到的翻译结果进一步逆翻成源语言的语句而显示于逆翻译区域 160 内,能够判定翻译是否正确传达了讲话者的意图。其中,在本实施方式的说明中,为了易于理解实施方式的说明,并不说明与该逆翻译相关联的功能部分的详细情况。

[0050] 在翻译结果显示区域 170 显示:目标语言的语言名 174;自动翻译结果的语句(目标语言的语句)的文本 176;以及用于再生文本 176 的合成声音的再生按钮 172。声音翻译的结果被自动作为合成声音而讲出,在想要反复听取的情况下操作再生按钮 172。

[0051] 在讲话候补推荐区域 190 显示:由被推荐为接着刚刚之前的用户的讲话而发言的可能性高且在自动翻译中受理的可能性高的讲话的讲话构成的讲话候补推荐列表 192;以及与讲话候补推荐列表 192 的各讲话对应地显示且用于产生针对各讲话的翻译请求的翻译请求按钮 194。

[0052] 在状态区域 180 显示利用次数等系统的利用状况和麦克按钮 182。麦克按钮 182 与声音输入按钮 152 同样地具有开始录音的功能,但是与声音输入按钮 152 不同,如果对麦克按钮 182 按压一次后放开则开始声音的录制,如果再次按压麦克按钮 182 后放开则结束声音的录制。

[0053] - 声音翻译的序列 -

[0054] 参照图 3,说明使用了声音翻译系统 100 的声音翻译时便携式终端 104 与服务器 106 之间的典型的通信序列。最初,在便携式终端 104 中进行声音输入 200,生成包含该声音、声音翻译的语言的组合等信息以及从传感器的集合得到的环境信息的声音识别请求 202。声音识别请求 202 从便携式终端 104 被发送到服务器 106。服务器 106 接收到该声音识别请求时进行声音识别处理 220,并输出声音识别结果的文本(在本实施方式中,声音识别处理 220 进行统计声音识别处理,假设伴随识别的得分最高的仅有 1 个)。作为输入,将该文本提供给自动翻译处理 222。自动翻译处理 222 将输入的源语言的语句自动翻译成目标语言,并生成目标语言的语句。该目标语言的语句被提供至声音合成处理 224。声音合成处理 224 根据所提供的目标语言的语句,合成声音。

[0055] 另一方面,声音识别处理 220 的识别结果的文本也与环境信息一起被提供至讲话候补列表生成处理 230。服务器 106 包括:讲话连续模型 226,是以能够针对讲话集合内的各讲话计算出在所提供的讲话之后讲出的概率的方式,预先学习完毕的统计模型;以及讲话候补存储部 228,用于存储源语言的讲话的、附加了表示自动翻译处理 222 的受理容易度的翻译得分的讲话的集合。在讲话候补列表生成处理 230 中,使用通过讲话连续模型 226 针对讲话集合内的各讲话计算出的概率和存储在讲话候补存储部 228 中的每个讲话的翻译得分,选择规定个数在下次被讲出的可能性高且翻译得分高的讲话,作为列表来输出。本实施方式中,在该评价中,使用由某一讲话在下次被讲出的概率与该讲话的翻译得分之

间的线性和构成的评价得分。在该实施方式的情况下,将评价得分高的讲话设为适于下次候补的讲话。因此,该线性和的各系数都为正。该系数的值随着概率的次序和翻译得分的次序的不同而不同,所以需要基于具体的组合来适当地确定。

[0056] 讲话连续模型 226 以如下方式进行了统计完成学习:如果接受了讲话的文本和进行该讲话时的环境信息,则能够针对规定的讲话集合之中的讲话,计算出在由文本表示的讲话之后连续讲出的概率。

[0057] 声音识别处理 220 输出的声音识别结果的文本、通过自动翻译处理 222 得到的翻译结果的文本、通过声音合成处理 224 合成的合成声音数据、以及讲话候补列表生成处理 230 生成的讲话候补列表都被提供至将这些数据发送到便携式终端 104 的发送处理部 232。发送处理部 232 将所提供的这些数据变换成用于发送的规定的格式,并发送到便携式终端 104。

[0058] 接收到来自发送处理部 232 的数据的便携式终端 104 显示接收到的声音识别结果、自动翻译结果以及讲话候补列表(步骤 204)。便携式终端 104 进一步地进行从发送处理部 232 接收到的合成声音的讲话(步骤 206)。在讲话后,便携式终端 104 返回等待接受下一次声音输入 200 的状态。

[0059] 另外,如前所述,图 3 示出的是典型的处理序列。在不是采用声音输入而是进行了文本输入的情况下,执行与图 3 所示的处理序列不同的处理序列。

[0060] - 便携式终端 104 -

[0061] 参照图 4,便携式终端 104 包括:处理器 250,其通过执行规定的程序来控制便携式终端 104 的各部分,从而实现各种功能;存储器 252,其存储处理器 250 执行的程序以及该程序的执行所需的数据,也起到处理器 250 的作业区域的作用;以及处理器 250 与后述的各种传感器等之间的接口 254。以下说明的结构要素都能够经由接口 254 而与处理器 250 进行通信。

[0062] 便携式终端 104 还包括:GPS 接收机 258,用于通过 GPS 功能来获取便携式终端 104 的位置的经度以及纬度信息;用于检测便携式终端 104 的 3 轴方向的加速度的加速度传感器 260、用于检测与便携式终端 104 的 3 轴相关的倾斜度的倾斜度传感器 262、检测便携式终端 104 周围的磁性的磁性传感器 264、检测便携式终端 104 周围的明亮度的明亮度传感器 266、探测便携式终端 104 的规定位置涉及的压力的压力传感器 268 以及检测便携式终端 104 周围的温度的温度传感器 270 等多个传感器;麦克风 256;能够通过无线通信经由未图示的基站与因特网 102 连接的通信装置 272;触摸面板 274;与触摸面板 274 分开设置于便携式终端 104 的筐体上的操作按钮 276;以及扬声器 280。

[0063] GPS 接收机 258、加速度传感器 260、倾斜度传感器 262、磁性传感器 264、明亮度传感器 266、压力传感器 268 以及温度传感器 270 等用于收集表示讲话时的环境的信息。在本实施方式中,声音输入时的这些各种传感器的输出与声音翻译的语言对等设定信息和从声音得到的 ADPCM 声音信号一起以规定的形式被发送到服务器 106,作为表示讲话时的环境的环境信息。这样发送到服务器 106 的信息是声音翻译请求,该声音翻译请求请求针对讲话声音的声音识别、和针对声音识别的结果的数据处理即自动翻译以及翻译结果的声音合成。

[0064] 实现便携式终端 104 的功能的各种程序之中,用于利用声音翻译服务的应用程序

具有如下的控制构造。参照图 5,如果启动该程序,则进行存储器区域的确保、以规定的初始值来初始化各存储器位置的初始设定处理(步骤 300)。在初始化完成后,在便携式终端 104 的触摸面板 274,显示用于声音翻译服务的初始画面(步骤 302)。在初始画面中,激活声音输入按钮 152、文本输入按钮 154、麦克按钮 182 以及设定按钮 142,使翻译按钮 164 以及再生按钮 172 失效。

[0065] 接着,等待来自用户的输入,根据进行了什么样的输入来使控制的流程分支(步骤 304)。

[0066] 如果按压了声音输入按钮(图 2 的声音输入按钮 152),则执行声音输入处理(步骤 310)。声音输入处理通过调取声音输入的 API(Application Programming Interface)来进行。接着,对输入的声音进行规定的信号处理,生成 ADPCM(Adaptive Differential Pulse Code Modulation)形式的声音信号(步骤 312)。进一步,基于该声音信号、此时的各传感器的输出以及翻译语言等设定信息,生成声音翻译请求,并发送给服务器 106(步骤 314)。之后,从服务器 106 接收声音识别结果、自动翻译结果、其合成声音、讲话候补的列表(步骤 316),为了将声音识别结果的文本、自动翻译结果的文本以及讲话候补的列表分别显示于图 2 的输入文本显示区域 150、逆翻译区域 160 以及翻译结果显示区域 170,保存在规定的存储器区域中(步骤 318)。进一步,驱动扬声器 280 来产生自动翻译结果的合成声音(步骤 320)。即,通过驱动扬声器 280,以声音的形式对用户提示所请求的讲话的翻译结果。最后,进行应用程序画面 130 的更新(步骤 322),返回到步骤 304 的输入等待状态。此时,除了激活声音输入按钮 152、文本输入按钮 154 以及麦克按钮 182 以外,还激活再生按钮 172。

[0067] 如果在步骤 304 中按压了文本输入按钮(图 2 的文本输入按钮 154),则通过调取文本输入的 API,接受文本的输入(步骤 340),保存所输入的文本(步骤 342),并将所输入的文本显示于输入文本显示区域 150 以及逆翻译区域 160,由此更新画面(步骤 322),然后返回到步骤 304。此时,在更新后的画面中激活翻译按钮 164,并使再生按钮 172 失效。

[0068] 如果在步骤 304 中按压了翻译按钮(图 2 的翻译按钮 164),则使用在步骤 340 中输入并在步骤 342 中保存的文本、此时的各种传感器的输出、以及设定信息,来生成文本翻译请求,并发送到服务器 106(步骤 360)。接着,接收针对该请求的翻译结果、翻译结果的合成声音以及讲话候补推荐列表(步骤 362)。将接收到的翻译结果的文本显示于翻译结果显示区域 170(图 2)(步骤 364),并讲出翻译结果的合成声音(步骤 366)。之后,更新应用程序画面 130(步骤 322),返回到步骤 304。

[0069] 最后,如果操作了图 2 所示的设定按钮 142,则通过显示为了进行设定变更而预先准备的画面来受理设定的变更(步骤 380),如果设定结束,则将该设定结果保存到存储器 252 中(步骤 382),并按照变更后的设定来更新应用程序画面 130(步骤 322),将控制返回到步骤 304。

[0070] - 服务器 106-

[0071] 参照图 6,服务器 106 具有如下的功能性结构。另外,服务器 106 通过作为硬件而具有大容量的外部存储装置的计算机和在该计算机上执行的声音翻译服务器程序实现。

[0072] 服务器 106 包括:接收处理部 406,用于经由因特网 102(参照图 1),接收来自便携式终端 104 这样的不确定的多个终端的声音翻译请求以及文本翻译请求;以及控制部 408,

用于分析接收处理部 406 接收到的请求,并通过按照分析结果将数据提供给规定的功能模块或者输出用于数据选择的控制信号,从而进行声音翻译的整体控制。服务器 106 还包括:按语言分类的声音识别用资源 400,是为了进行多个源语言的声音识别以及为了进行自动翻译中的目标语言语句的生成而预先准备的,且是针对多个语言分别准备的;按语言对分类的资源 402,是为了进行自动翻译而按多个源语言和多个目标语言的组合的每一个组合而预先准备的;以及按语言分类的声音合成用资源 404,是为了进行声音合成而针对各个目标语言预先准备的。

[0073] 在本实施方式中,按语言分类的声音识别用资源 400 包括用于声音识别的各语言的统计音响模型、辞典以及统计语言模型。在语言成为目标语言时,在自动翻译中也使用辞典以及语言模型。因此,服务器 106 还包括:选择部 410,从按语言分类的声音识别用资源 400 之中,选择由声音翻译请求指定的源语言的资源;以及选择部 412,从按语言分类的声音识别用资源 400 之中,选择由声音翻译请求指定的目标语言的资源。

[0074] 按语言对分类的资源 402 包括统计翻译模型,按源语言和目标语言的每一个组合预先准备该统计翻译模型。服务器 106 还包括选择部 414,该选择部 414 从按语言对分类的资源 402 之中选择由声音翻译请求指定的语言对的资源。

[0075] 按语言分类的声音合成用资源 404 包括进行目标语言的声音合成时所需的按语言分的资源。如果声音合成是原片连接型,则按语言分类的原片 DB 包含在该资源中。服务器 106 还包括选择部 416,该选择部 416 从按语言分类的声音合成用资源 404 之中选择由声音翻译请求指定的目标语言的资源。

[0076] 控制部 408 具有以下功能:基于包含在声音翻译请求中的源语言和目标语言的组合,对选择部 410、选择部 412、选择部 414 以及选择部 416,分别发送控制信号以便选择适当的资源。

[0077] 服务器 106 还包括:声音识别引擎 418,从控制部 408 接受声音翻译请求中的 ADPCM 数据,使用由选择部 410 选择出的资源,进行源语言的声音识别并输出其文本数据;自动翻译引擎 422,接受声音识别引擎 418 输出的源语言的文本数据,使用由选择部 412 选择出的目标语言的资源、和由选择部 414 选择出的与源语言和目标语言的语言对相对应的资源,将源语言语句翻译成目标语言语句,并作为文本数据来输出;以及声音合成处理部 424,用于接受自动翻译引擎 422 输出的目标语言的文本数据,使用由选择部 416 选择出的语言资源,进行声音合成。

[0078] 在声音识别引擎 418 的输出和自动翻译引擎 422 的输入之间插入具有 2 个输入和与自动翻译引擎 422 的输入连接的输出的选择部 420。在选择部 420 的一个输入上连接声音识别引擎 418 的输出 438。向另一个输入输入从控制部 408 输出的文本翻译请求中的文本数据 440。如前所述,在便携式终端 104 中,不仅采用声音输入,还进行文本的输入。在声音输入的情况下,将作为声音识别引擎 418 的输出 438 的文本数据提供至自动翻译引擎 422,在文本输入的情况下,不经由声音识别引擎 418,而是直接将请求中的文本数据 440 提供至自动翻译引擎 422。选择部 420 选择 2 个输入中的哪一个是控制部 408 查看声音翻译请求的内容后通过切换信号 442 来控制的。针对文本翻译请求的处理仅是输入与针对声音翻译请求的处理不同,自动翻译以后的处理与声音翻译请求的处理相同。因此,在此不重复针对文本翻译请求的处理的详细的说明。另外,对声音识别引擎 418 输出的文本数据、对从

控制部 408 直接提供至选择部 420 的文本数据都赋予声音翻译请求中的各种传感器的输出值。

[0079] 服务器 106 还包括:为了将附加在声音翻译请求中的各种传感器的输出值变换成在推测下一次讲话候补时使用的特征量向量的要素值而使用的要素决定用各种 DB430;针对成为对象的多个语言准备的多个讲话连续模型 226 的存储装置;针对成为对象的多个语言准备的多个讲话候补存储部 228;以及选择部 434 及选择部 436,用于按照控制部 408 的控制,从多个讲话连续模型 226 以及多个讲话候补存储部 228 中选择与讲话语言相对应的模型以及讲话候补存储部,且该选择部 434 以及选择部 436 与讲话概率计算部 426 以及讲话候补推荐列表生成部 428 连接。服务器 106 还包括:讲话概率计算部 426,用于接受选择部 420 输出的文本数据,使用要素决定用各种 DB430 以及讲话连续模型 226,针对预先准备的各个讲话计算出成为所提供的讲话的下一讲出的概率;讲话候补推荐列表生成部 428,用于基于由讲话概率计算部 426 针对各讲话计算出的概率、和存储在讲话候补存储部 228 中的各讲话的翻译以及声音识别得分,从该得分的上位中选择多个下一次被讲出的可能性高且由自动翻译引擎 422 受理(被正确地翻译)的可能性高的讲话,由此生成讲话候补推荐列表;以及发送处理部 432,用于根据自动翻译引擎 422 输出的翻译结果的文本数据、声音合成处理部 424 输出的合成声音、以及讲话候补推荐列表生成部 428 生成的讲话候补推荐列表,来组成规定数据格式的反馈数据,向发送了声音翻译请求的终端(便携式终端 104 等)进行反馈。

[0080] 但是,对于讲话连续模型 226 以及讲话候补存储部 228 而言,需要预先准备。为此,设置有模型生成部 108。针对模型生成部 108 的结构将后述。

[0081] 参照图 7,用于通过由构成服务器 106 的计算机的硬件执行从而实现图 6 所示的控制部 408 的功能的程序具有如下的控制构造。如果启动该程序,则包括:首先进行必要的存储区域的确保以及初始化等在启动之后需要立即执行 1 次的处理的初始设定步骤 450;在初始设定后,等待经由因特网 102 从其他的终端装置接收与声音翻译相关的请求的步骤 452;以及根据在步骤 452 中接收到的请求是什么来使控制的流程分支的步骤 454。

[0082] 如果判定为请求是声音翻译请求,则基于包含在请求中的语言对的信息,设定源语言和目标语言,并设定图 6 所示的选择部 410、选择部 412、选择部 414 以及选择部 416(步骤 460),执行声音识别(步骤 462)。如果声音识别结束,则将结果的源语言的文本作为输入,进行向目标语言的自动翻译(步骤 464)。如果自动翻译结束,则将通过自动翻译得到的输出的文本作为输入,进行声音合成(步骤 466)。进一步,基于在步骤 462 中得到的文本数据,生成用于推测下一次讲话的特征量向量(步骤 468)。在生成该特征量向量时,有时使用包含在请求中的原始的传感器输出,有时使用要素决定用各种 DB430(参照图 6)将原始的传感器输出变换为其他范畴的值。例如,虽然 GPS 的输出是纬度及经度的信息,但是可以将这些信息直接作为特征量向量的要素,也可以将纬度以及经度的信息与存在于该位置的设施名或包含该位置的地域名之间的对应关系存储在要素决定用各种 DB430 中,并在变换为设施名或地域名等之后,将变换后的值作为特征量向量的要素。利用所生成的特征量向量,使用图 6 所示的讲话概率计算部 426,推测规定个数的下一次讲出的概率高的讲话候补,进一步使用图 6 所示的讲话候补存储部 228,根据将各讲话候补的概率和它们的翻译以及声音识别得分组合而得到的评价得分,选择规定个数的成为下一次讲话的概率高且翻译

得分也高的讲话候补,由此生成讲话候补推荐列表(步骤470)。最后,向对方终端发送在步骤462中得到的源语言的声音识别结果、在步骤464中得到的自动翻译结果的文本、在步骤466中得到的声音合成、在步骤470中得到的讲话候补推荐列表(步骤472),并将控制返回到步骤452。

[0083] 另一方面,如果在步骤454中判定为请求是文本翻译请求,则根据包含在请求中的语言对的设定信息,设定图6所示的选择部410、选择部412、选择部414以及选择部416(步骤478),将所输入的源语言的文本自动翻译成目标语言(步骤480)。基于得到的目标语言的文本,对目标语言的声音进行合成(步骤482),并基于包含在请求中的源语言的输入文本和环境信息,生成特征量向量(步骤484),通过参照讲话连续模型226,从而生成由下一次讲出的可能性高且翻译得分也高的讲话候补构成的讲话候补推荐列表(步骤486)。最后,将在步骤480中得到的目标语言的文本数据、在步骤482中得到的目标语言的合成声音、和在步骤486中得到的讲话候补发送到便携式终端(步骤488),并将控制返回到步骤452。

[0084] 以上是用于实现图6所示的服务器106的程序的构造的概况。

[0085] 需要预先准备图6所示的讲话连续模型226以及存储在讲话候补存储部228中的讲话候补的集合。图8示出用于此的模型生成部108的结构。另外,在本实施方式中,作为图6所示的要素决定用各种DB430,包括:GPS/地域信息变换DB518,存储了从GPS得到的纬度/经度信息、与由该纬度/经度信息确定的国家、地域、州、都道府县、市、镇等地域信息之间的对应关系;以及IP地址/设施名变换地址DB522,存储了IP地址、与分配给该IP地址的设施名之间的对应关系。

[0086] 参照图8,模型生成部108包括含有多个单个语言的讲话文本的文集510。存储在文集510中的讲话包括该讲话的声音数据和讲话的起始的文本。起始文本全都被分割成词素等规定的单位,对各单位附加了被称为DA(Dialog Activity)标签的表示进行讲话时的状况的标签。对各讲话进一步附加进行该讲话时的日期时间、讲话的用户的用户ID或终端ID、收集了该讲话时的用户的位置(通过GPS得到的纬度以及经度)、发送来该讲话的终端的IP地址、终端的各传感器探测到的加速度、倾斜度、磁性、明亮度、压力、温度等信息。这些信息之中,在因没有对应的传感器等情况而无法通过便携式终端得到的信息中代入表示没有信息的规定值。

[0087] 模型生成部108还包括:输入部512,在对包含在文集510中的讲话数据通过手动进行词素分析、标签附加等处理时,由用户使用该输入部512;以及基础特征向量生成部514,根据存储在文集510中的各语句,生成用于生成进行讲话连续模型226的学习用的学习数据的基础特征向量。基础特征向量生成部514至少针对存储在文集510中的各语句,以规定的顺序排列附加给这些语句的日期时间、用户ID或终端ID、以及其他各种传感器信息,进一步生成将下一次讲出的讲话的识别编号作为要素的向量。

[0088] 模型生成部108还包括:地域信息附加部516,其进行如下处理,即,将包含在由基础特征向量生成部514生成的各特征向量中的纬度/经度信息、与GPS/地域信息变换DB518进行对照,从而得到与该特征向量相对应的讲话完成的国家、地域、州、都府县、市、镇等地域信息,并插入到特征向量中的适当的位置;设施信息附加部520,进行如下处理,即,从地域信息附加部516接受特征向量,通过将其中包含的IP地址与IP地址/设施名变换

DB522 进行对照,从而得到完成该讲话的设施名,并插入到特征向量内的适当的位置;特征向量存储部 526,用于蓄存从设施信息附加部 520 输出的特征向量;以及讲话连续模型学习部 524,用于将存储在特征向量存储部 526 中的特征向量作为学习数据,进行讲话连续模型 226 的统计学习。

[0089] 模型生成部 108 还包括:讲话汇总部 540,用于对包含在文集 510 中的讲话之中相同的讲话进行汇总,从而生成仅由互不相同的讲话构成的集合;翻译引擎 544,将由讲话汇总部 540 汇总的各个讲话翻译成多个语言,并针对各讲话输出翻译结果的得分;以及翻译得分计算部 542,按每个讲话对通过由翻译引擎 544 翻译成多个语言而得到的翻译结果的得分求取平均,计算出各讲话的的平均的翻译得分,并存储至讲话候补存储部 228 中。翻译引擎 544 被假设为进行统计自动翻译,并将翻译结果的似然性设为翻译得分。可认为翻译得分越高,成为其根源的源语言的讲话越容易翻译。

[0090] 模型生成部 108 还包括:声音识别引擎 546,对包含在文集 510 中的各讲话的声音数据进行声音识别;以及声音识别得分计算部 548,用于对包含在文集 510 中的各讲话的起始数据、和针对该讲话的声音识别引擎 546 的识别结果进行比较,从而计算出各讲话的声音识别得分,附加到各讲话中并存储于讲话候补存储部 228。

[0091] 另外,在图 8 中,作为单一的部分而示出了模型生成部 108,但是模型生成部 108 需要按翻译的每个源语言来生成。关于翻译引擎 544,可以针对某源语言,使用所有可利用的翻译引擎,并对由它们得出的翻译得分求取平均,也可以仅将特定的多个语言作为目标语言来计算出翻译得分,并使用它们的平均。也可以根据情况,使用仅将某单一语言设为目标语言的翻译引擎 544。

[0092] < 动作 >

[0093] - 概要 -

[0094] 该声音翻译系统 100 按照以下方式进行动作。服务器 106 具有两个动作阶段。第 1 是基于模型生成部 108 的讲话连续模型 226 以及讲话候补存储部 228 的学习,第 2 是使用了学习结束的讲话连续模型 226 以及讲话候补存储部 228 的声音翻译服务的执行。以下,首先说明学习阶段的服务器 106 的动作,接着说明声音翻译服务阶段的便携式终端 104 和服务器 106 的动作。

[0095] - 学习 -

[0096] 需要预先进行讲话连续模型 226、以及存储在讲话候补存储部 228 中的各讲话的翻译及声音识别得分的学习。为此,首先,针对成为处理对象的所有语言,按每个语言分别汇集语句来生成文集 510。针对文集 510 内的各语句,预先进行词素分析等,虽然如果进行 DA 标签的赋予等是优选的,但是根据需要,使用输入部 512 进行这样的处理。

[0097] 针对包含在某语言的文集 510(包括声音数据和起始文本这两者)中的各语句,进行以下处理。即,使用附加给该语句的信息,通过基础特征向量生成部 514(图 8)生成基础的特征向量。此时,基于附加给各讲话的用户 ID 以及讲话的日期时间的信息,确定在某讲话的下次讲出哪一个讲话,将确定下次讲话的信息代入特征向量内的适当的位置的要素中。接着,通过将包含在各特征向量中的纬度/经度信息与 GPS/地域信息变换 DB518 进行对照,从而根据纬度/经度信息,得到国家、地域、州、都府县、市、镇等地域信息,并代入特征向量中的适当的要素中。在没有对应的信息的情况下,将表示没有该信息的情况的特

定值代入该要素（针对其他要素也是同样的）。进一步，通过将包含在特征向量中的 IP 地址与 IP 地址 / 设施名变换 DB522 进行对照，从而将收录了该讲话的 IP 地址变换成与该 IP 地址对应的设施名，并代入特征向量的适当的要素中。这样，将特征向量存储至特征向量存储部 526 中。

[0098] 如果完成向特征向量存储部 526 存储特征向量，则讲话连续模型学习部 524 执行讲话连续模型 226 的统计学习处理，或者与存储并行地执行讲话连续模型 226 的统计学习处理。通过该统计学习，能够得到如下的讲话连续模型 226：如果给出以讲话日期时间、用户 ID、IP 地址、纬度 / 经度信息、各种传感器的值作为要素的某讲话的特征向量，则能够针对各讲话计算出在该讲话的下次讲出的概率。这些讲话连续模型 226 优选存储至非易失性的存储装置中。

[0099] 另一方面，讲话汇总部 540 汇总包含在文集 510 中的各讲话之中的相同的讲话，作为文本。翻译得分计算部 542 通过将汇总后的各讲话提供至翻译引擎 544，从而计算出该讲话的翻译得分。此时，在本实施方式中，准备作为翻译引擎 544 来翻译到各目标语言的统计量，并将根据这些量得到的翻译结果的似然性的平均设为该讲话的翻译得分。翻译得分计算部 542 针对各源语言生成由各讲话得到的翻译得分构成的翻译得分 DB，并保存至讲话候补存储部 228。以上，讲话连续模型 226 以及翻译得分 DB 的学习结束。进一步地，通过声音识别引擎 546 以及声音识别得分计算部 548，计算出包含在文集 510 中的各语句的声音识别时的声音识别得分。即，声音识别引擎 546 进行针对各语句的声音数据的声音识别。声音识别得分计算部 548 对声音识别引擎 546 的声音识别结果和预先附加到该语句的起始语句进行比较，从而计算出该语句的声音识别得分。该声音识别得分按保存在文集 510 中的各讲话而存储至讲话候补存储部 228。其结果，能够使用存储在讲话候补存储部 228 中的信息，计算出存储在文集 510 中的各讲话的翻译以及声音识别得分。

[0100] - 声音翻译服务 -

[0101] 需要预先对便携式终端 104 等分配如图 2 所示的声音翻译应用程序。在本实施方式中，能够连接便携式终端 104 的服务器 106 是根据声音翻译应用程序来固定的。当然，如果服务器 106 有多个，则也可以由用户从其中选择期望的服务器。

[0102] 在想要利用服务器 106 的声音翻译服务的情况下，对用户来说大致存在 2 个选择分支。第 1 个是声音翻译，第 2 个是文本翻译。以下，首先说明利用声音翻译服务时的用户的操作以及便携式终端 104 及服务器 106 的动作，接着说明利用文本翻译时的用户的操作以及便携式终端 104 及服务器 106 的动作。另外，在此之前，用户需要通过操作图 2 的设定按钮 142 来调取设定画面，选择自己想要利用的源语言和目标语言的组合。

[0103] 在想要进行声音翻译的情况下，对用户来说存在 2 种方法。第 1 种是以下方法：按压声音输入按钮 152，并在按压期间讲话，讲话结束后放开声音输入按钮 152。第 2 种是：通过按压麦克按钮 182 来开始声音的录制并进行讲话，在讲话结束后通过再次按压麦克按钮 182 来结束声音的录制。不论选择哪一种方法，程序都能够在图 5 的步骤 310 以及步骤 312 中，通过调取与所选择的处理相对应的 API，来进行声音的录制和信号处理，生成规定形式的声音数据。

[0104] 如果声音的录制结束（如果放开声音输入按钮 152，或者在声音的录制执行中再次按压麦克按钮 182），则进行图 5 所示的步骤 314 的处理，将声音翻译的请求命令、基于设

定的声音对的信息、声音数据以及环境信息发送到服务器 106。在该请求中附加有基于设定的语言对的信息、讲话日期时间、用户的识别信息以及由 GPS 接收机 258、加速度传感器 260、倾斜度传感器 262、磁性传感器 264、明亮度传感器 266、压力传感器 268 以及温度传感器 270 的输出构成的环境信息。

[0105] 服务器 106 如果接收到该声音翻译请求（图 7 的步骤 452），则按照请求中的语言对信息来选择语言对（步骤 454 至步骤 460），并按照选择出的语言的组合来控制选择部 410、选择部 412、选择部 414、选择部 416、选择部 434 以及选择部 436，选择适当的要素。服务器 106 进一步进行声音识别（步骤 462）、针对声音识别结果的自动翻译（步骤 464）、针对翻译结果的声音合成（步骤 466）。之后，服务器 106 根据声音识别的结果和附加到声音翻译请求中的各种信息，生成输入讲话的特征向量（步骤 468），使用讲话连续模型 226 以及讲话候补存储部 228 内的各讲话的翻译以及声音识别得分，生成由下一次讲出的可能性高且翻译以及声音识别得分也高的讲话候补构成的讲话候补推荐列表（步骤 478）。服务器 106 在最后将声音识别结果、翻译结果的文本数据、其合成声音、讲话候补推荐列表发送到便携式终端 104（步骤 472）。如果该处理结束，则服务器 106 转移到下一次的请求处理（步骤 452）。

[0106] 参照图 2，便携式终端 104 如果受理来自服务器 106 的反馈（图 5 的步骤 316），则将声音识别结果显示于输入文本显示区域 150，将自动翻译结果的文本显示于翻译结果显示区域 170，并将讲话候补推荐列表 192 以及翻译请求按钮 194 显示于讲话候补推荐区域 190（步骤 318），进一步地通过按照合成声音数据来驱动扬声器 280，由此进行讲话（步骤 320）。根据步骤 318 的结果来更新画面（步骤 322），并返回到用户的操作等待状态（步骤 304）。

[0107] 在用户下一次讲话时，能够参考显示于讲话候补推荐区域 190 的讲话候补推荐列表。由于并不是在什么都没有的状态下考虑下一次的讲话，而是能够关注具体的语句进行下一次讲话，所以用户不会感到不知所措。进一步地，由于作为讲话，显示接着之前的讲话讲出的可能性高的讲话，所以用户在操作便携式终端 104 时应当回想的事项少的可能性高。并且，显示于讲话候补推荐列表的讲话由于翻译以及声音识别得分高，所以如果进行了学习该讲话的讲话，则在进行声音翻译时，声音识别的结果正确的可能性变高，进一步地，使用了该结果的自动翻译的结果作为目标语言的讲话而成为正确的讲话的可能性也变高。因此，能够无误解且顺畅地推进便携式终端 104 的用户和其他语言的用户之间的对话。

[0108] 在本实施方式中，显示于讲话候补推荐列表 192 的翻译请求按钮 194 都起到产生针对相对应的讲话的翻译请求的按钮的作用。即，如果用户按压与讲话候补中的任一个文本相对应的翻译请求按钮 194，则将该文本选择为下一次讲话并发送到服务器 106，作为下一次自动翻译的对象。

[0109] 在文本翻译的情况下，便携式终端 104 如以下方式进行动作。参照图 2，用户操作文本输入按钮 154，调取文本输入画面。在便携式终端 104 中，在图 5 所示的流程图中，调取步骤 304 至步骤 340。如果文本的输入结束，则用户按压结束文本输入的按钮（在图 2 中未图示）。便携式终端 104 保存所输入的文本（步骤 342），并更新画面，以便将所输入的文本显示于输入文本显示区域 150（步骤 322）。如果用户进一步按压翻译按钮 164，则在图 5 中执行步骤 304 ~ 360 的处理，生成文本翻译请求，并发送到服务器 106。在该请求中附加文

本翻译命令、与基于设定的语言对相关的信息、所输入的源语言的文本数据、按压翻译按钮 164 时的日期时间及用户的识别信息、以及各种传感器的输出。

[0110] 服务器 106 如果接收到该请求,则通过图 7 所示的步骤 454、478、480、482、484、486 的途经,将输入文本翻译成目标语言,并对其声音进行合成。服务器 106 进一步根据输入文本和附加到文本翻译请求中的各种信息,生成特征向量,使用存储在讲话连续模型 226 以及讲话候补存储部 228 中的各讲话的翻译得分,生成由作为下一次讲话的可能性高且翻译得分也高的源语言的讲话列表构成的讲话候补推荐列表。并且,向便携式终端 104 反馈翻译结果文本、其合成声音以及讲话候补推荐列表。

[0111] 便携式终端 104 如果接受了该反馈,则将翻译结果文本显示于翻译结果显示区域 170,并将讲话候补推荐列表 192 以及翻译请求按钮 194 显示于讲话候补推荐区域 190。

[0112] 由于在该情况下,也与声音翻译时同样地,将用户考虑下一次讲话用的讲话候补推荐列表显示于讲话候补推荐区域 190,所以即便是采用声音进行下一次讲话,还是进行按钮输入,都不会感到不知所措,能够容易考虑下一次的讲话。

[0113] 如以上所述,根据本实施方式,在进行声音翻译的声音翻译系统 100 中,在便携式终端 104 的用户讲话之后,考虑下一次讲话时,将讲话候补推荐列表显示于便携式终端 104 的画面上。被显示的讲话候补都是接着先前的讲话讲出的可能性高且在翻译时被正确翻译的可能性高的候补。因此,用户通过参考这些讲话候补来考虑下一次的讲话,能够顺畅地继续与对方的对话,并且能够在相互交流中减少因误译产生误解的可能性。

[0114] 本次公开的实施方式仅是例示,本发明并不仅限于上述实施方式。本发明的范围在参考发明的详细说明书的记载的基础上,由权利要求书的各项示出,包括与记载于权利要求书的用语同等的意义以及范围内的所有变更。

[0115] 工业上的可利用性

[0116] 本发明能够利用于对多人特别是以不同的语言作为母语的人之间的交流进行辅助的终端装置。

[0117] 符号说明

[0118] 100 声音翻译系统

[0119] 104 便携式终端

[0120] 106 服务器

[0121] 108 模型生成部

[0122] 130 应用程序画面

[0123] 140 语言显示区域

[0124] 150 输入文本显示区域

[0125] 160 逆翻译区域

[0126] 180 状态区域

[0127] 190 讲话候补推荐区域

[0128] 192 讲话候补推荐列表

[0129] 220 声音识别处理

[0130] 222 自动翻译处理

[0131] 224 声音合成处理

- [0132] 226 讲话连续模型
- [0133] 228 讲话候补存储部
- [0134] 230 讲话候补列表生成处理
- [0135] 232 发送处理部
- [0136] 418 声音识别引擎
- [0137] 422 自动翻译引擎
- [0138] 424 声音合成处理部
- [0139] 426 讲话概率计算部
- [0140] 428 讲话候补推荐列表生成部

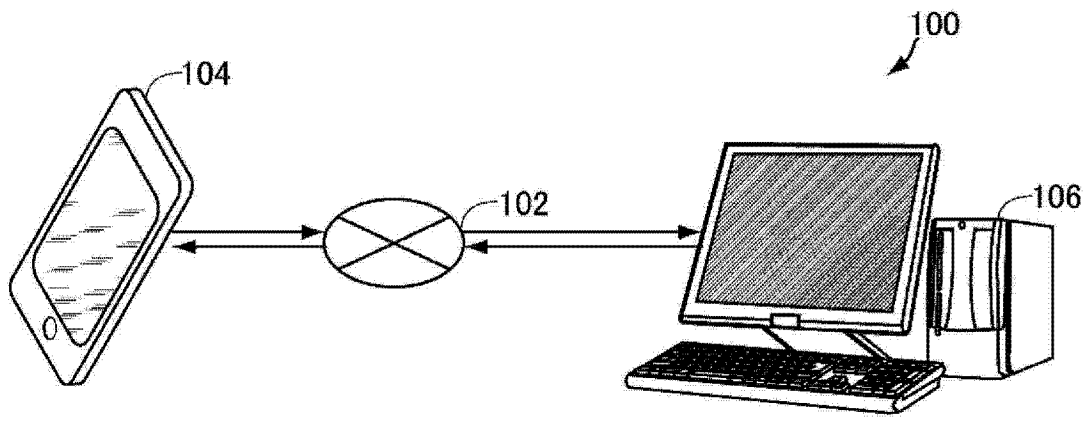


图 1

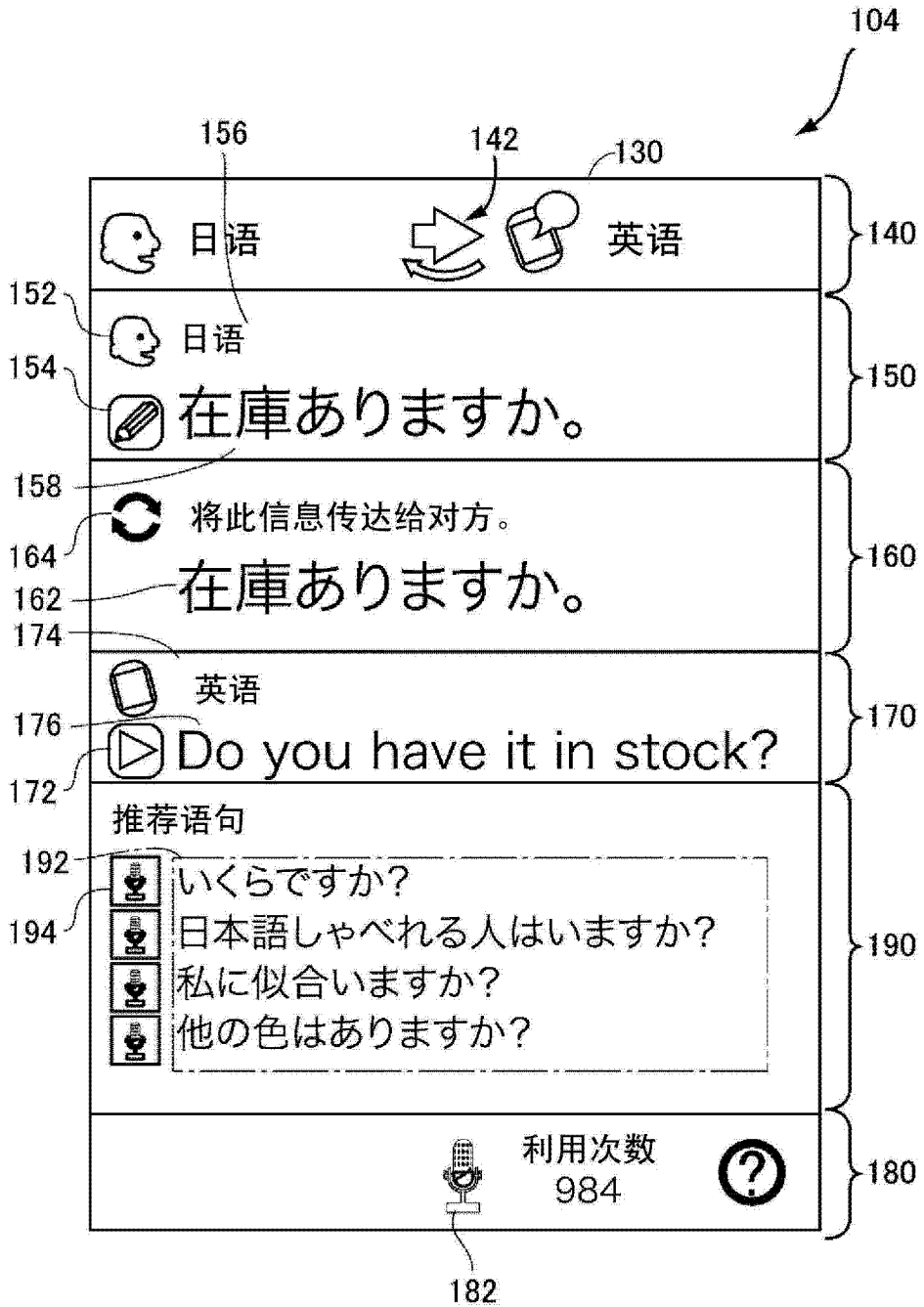


图 2

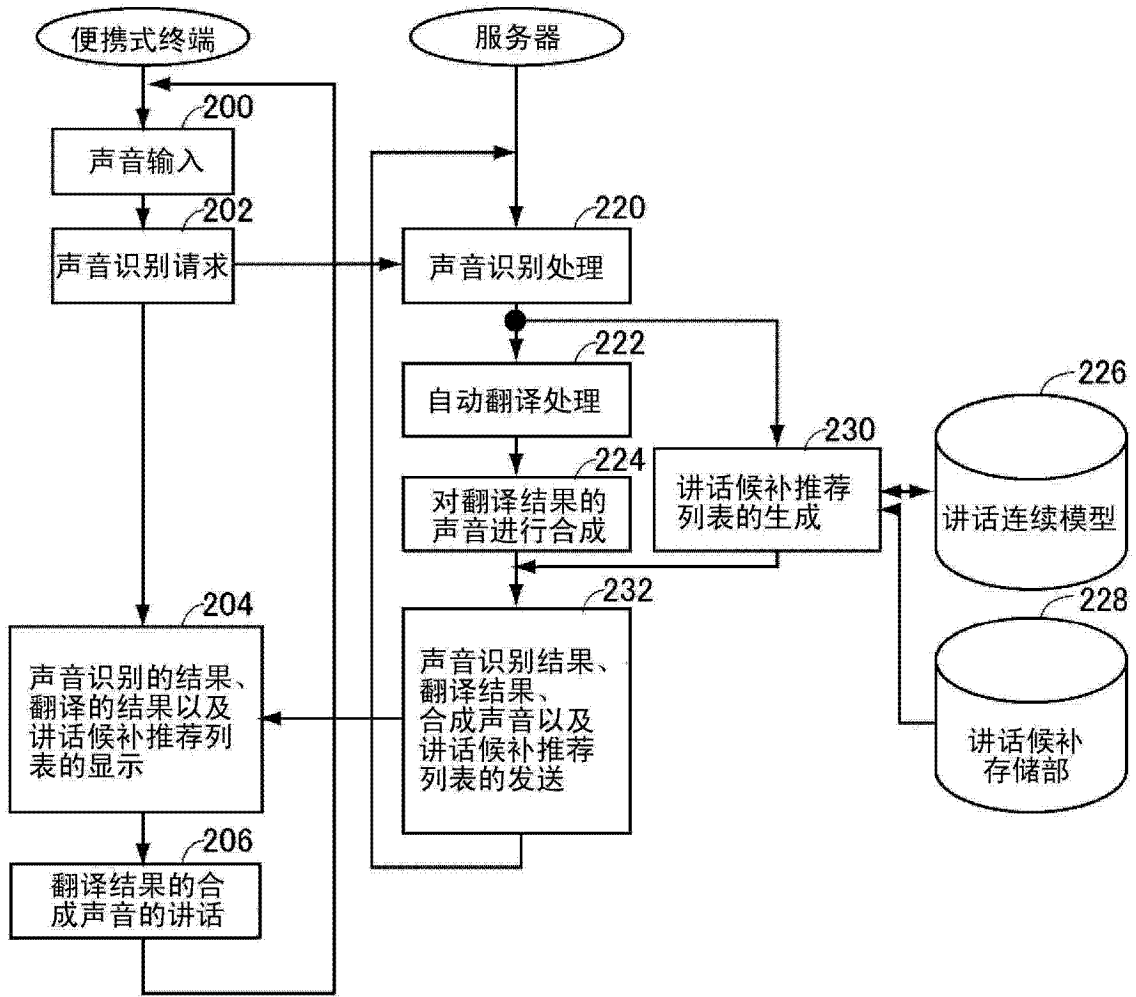


图 3

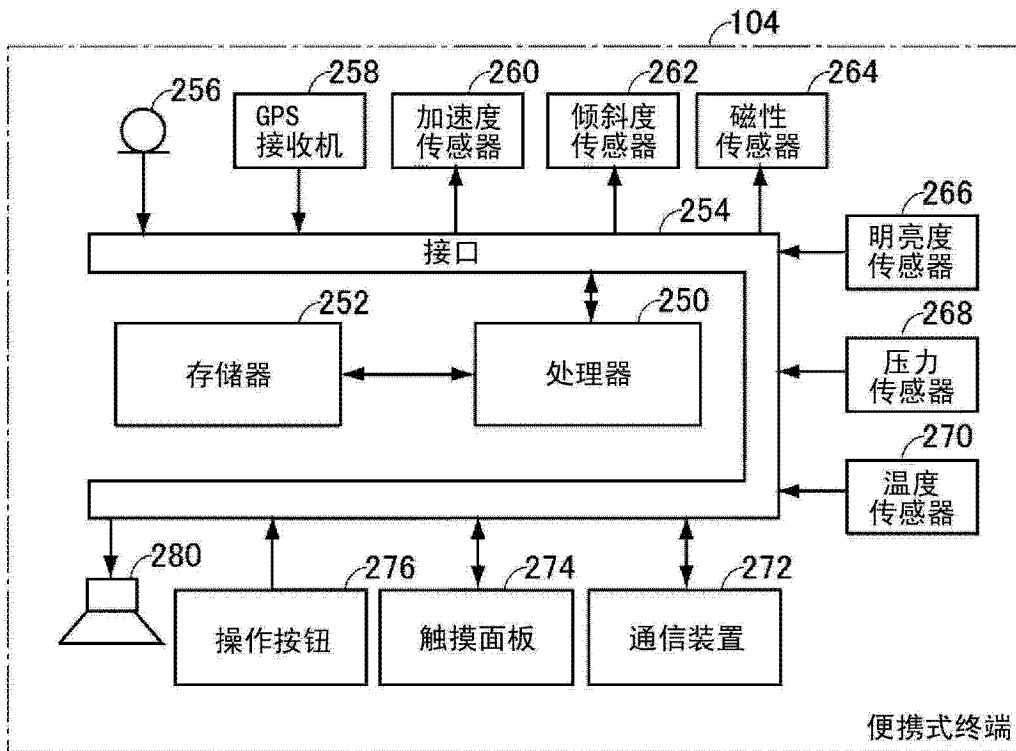


图 4

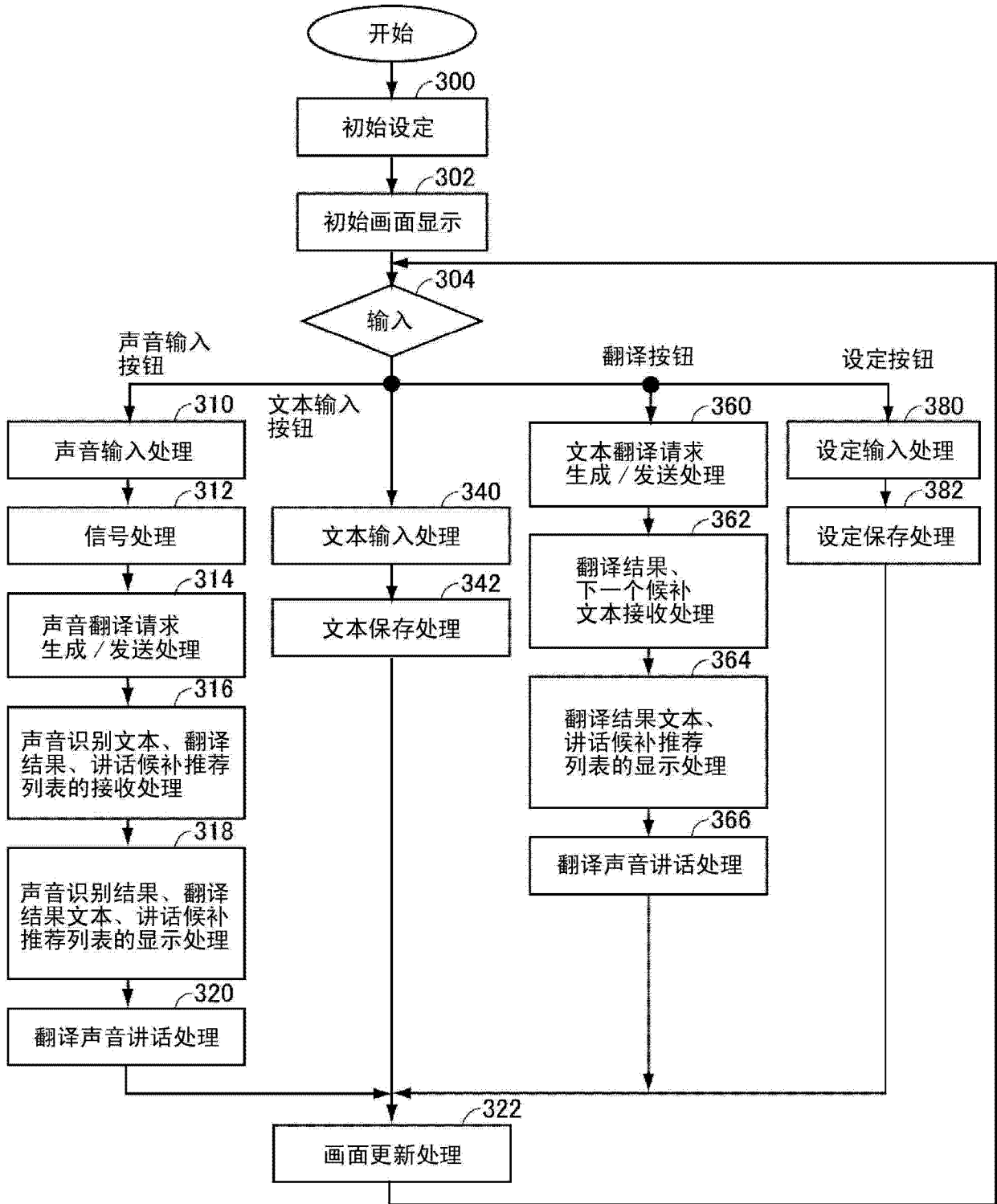


图 5

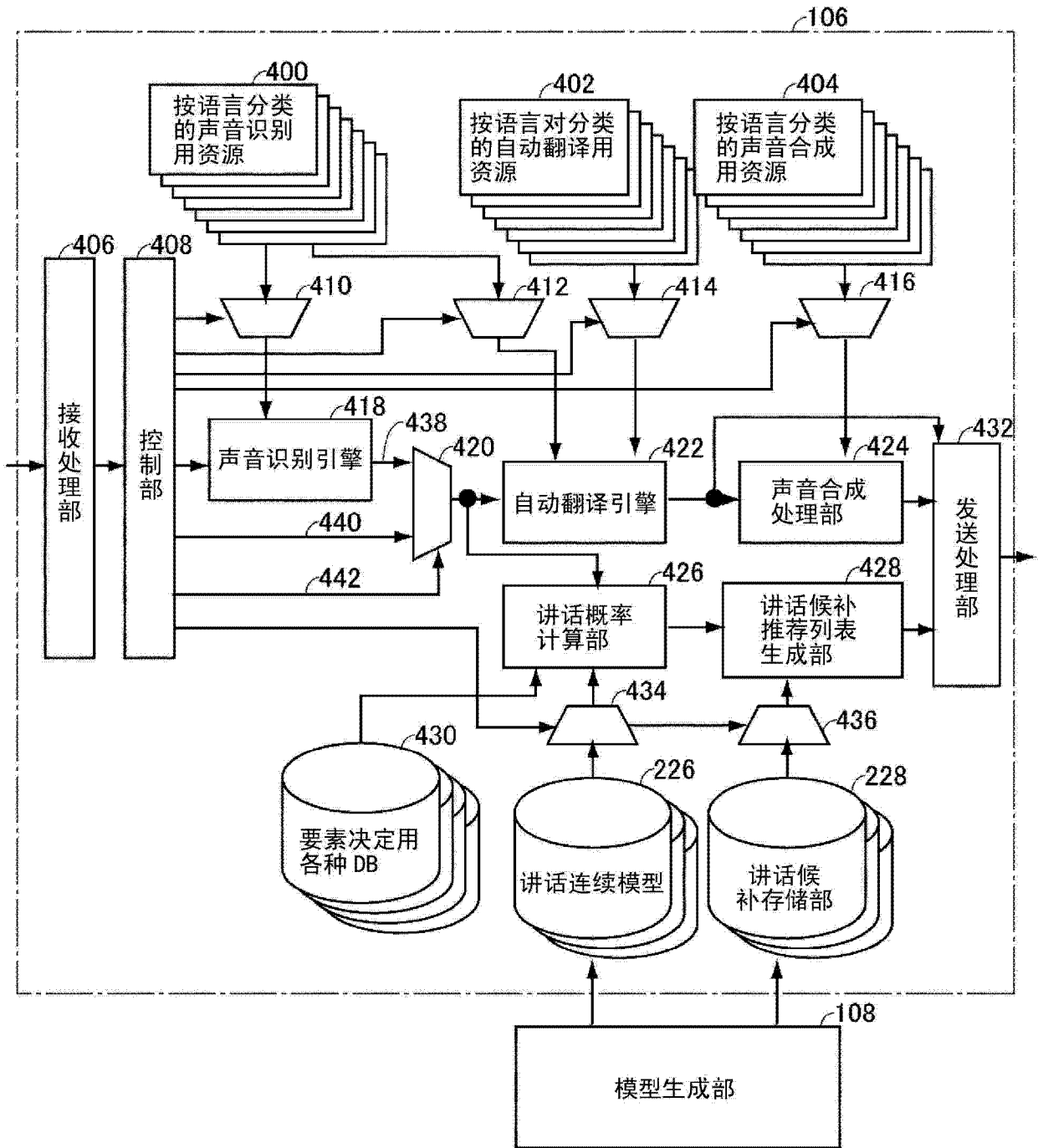


图 6

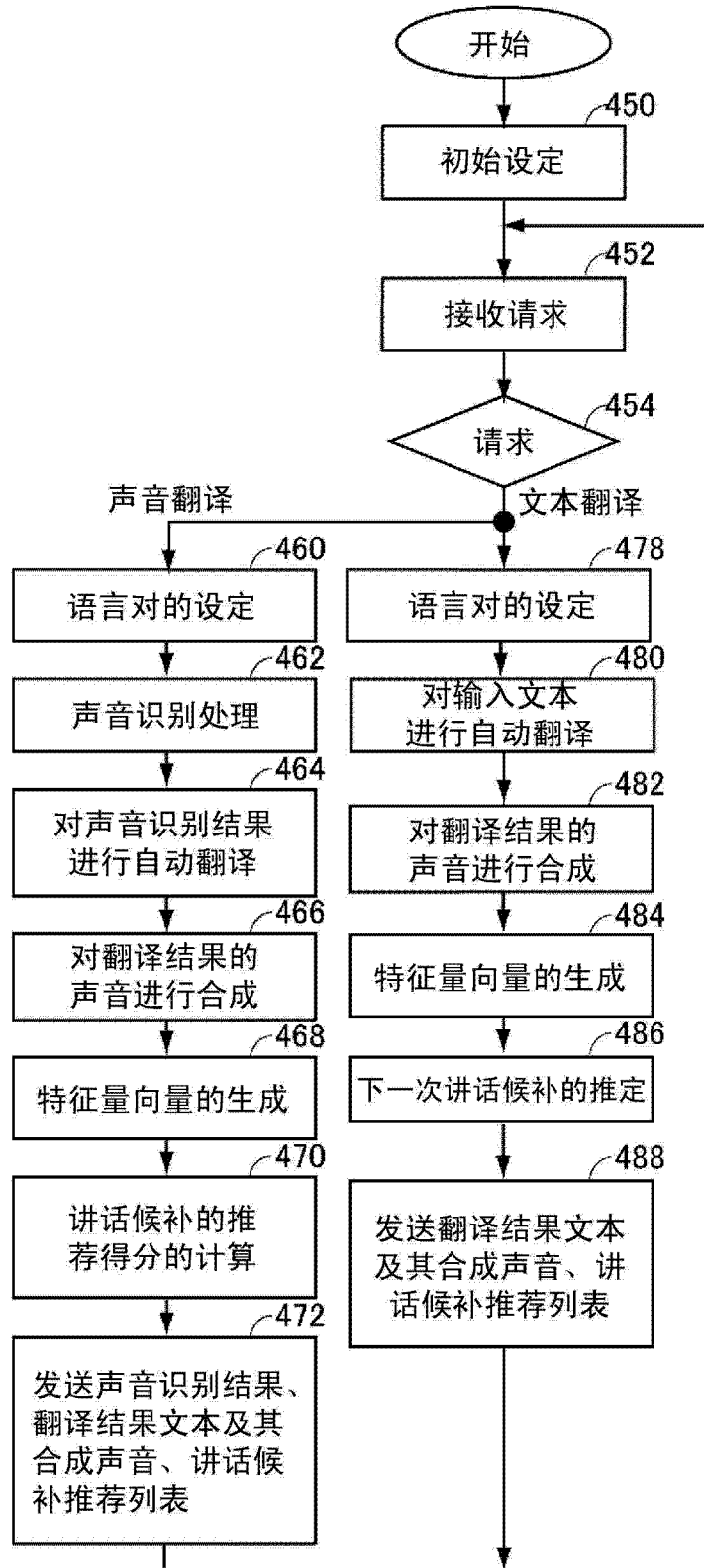


图 7

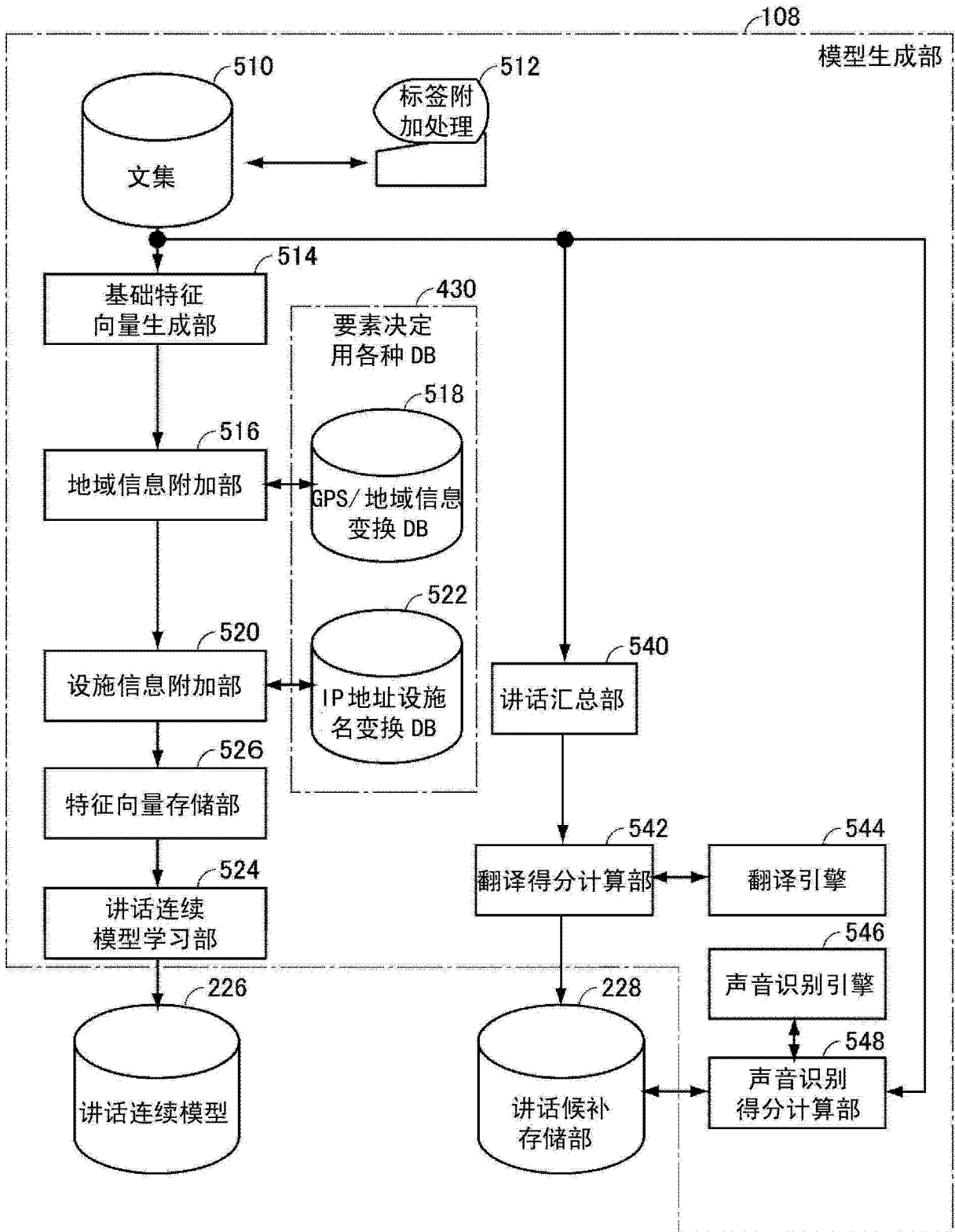


图 8