



(12) 发明专利

(10) 授权公告号 CN 109065016 B

(45) 授权公告日 2021.04.13

(21) 申请号 201811004153.8

审查员 赵磊

(22) 申请日 2018.08.30

(65) 同一申请的已公布的文献号

申请公布号 CN 109065016 A

(43) 申请公布日 2018.12.21

(73) 专利权人 出门问问信息科技有限公司

地址 100007 北京市海淀区中关村大街19号办公A楼10层1001

(72) 发明人 叶顺平

(74) 专利代理机构 北京睿派知识产权代理事务

所(普通合伙) 11597

代理人 刘锋

(51) Int. Cl.

G10L 13/02 (2013.01)

G10L 13/04 (2013.01)

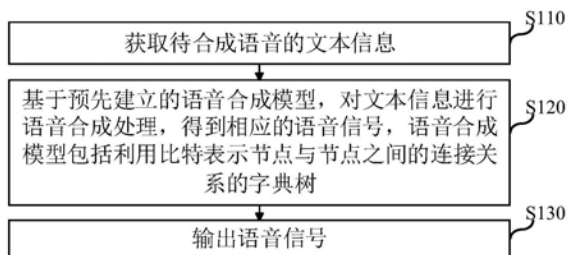
权利要求书2页 说明书11页 附图2页

(54) 发明名称

语音合成方法、装置、电子设备及非暂态计算机存储介质

(57) 摘要

本发明实施例涉及语音处理技术领域,提供了一种语音合成方法、装置、电子设备及非暂态计算机存储介质,其中,语音合成方法包括:获取待合成语音的文本信息;接着基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树;接着输出语音信号。本发明实施例的方法,利用比特表示节点与节点之间的连接关系的字典树替代传统的利用指针表示节点与节点之间的连接关系的字典树,极大减小了占用的内存空间,利于后续提高内存空间的使用效率。



1. 一种语音合成方法,其特征在于,包括:

获取待合成语音的文本信息;

对所述文本信息进行分词,得到至少一个词语;基于各个词语在文本信息中的逻辑顺序,根据利用比特表示节点与节点之间的连接关系的字典树,依次确定各个词语中各个字分别对应的拼音,所述字典树的节点包括字的节点以及词语的节点,词语的节点中携带相应词语中各个字分别对应的拼音;将确定得到的各个词语中各个字分别对应的拼音,依据各个词语在文本信息中的逻辑顺序进行缓存,得到相应的拼音序列;

对所述拼音序列进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树;

输出所述语音信号。

2. 根据权利要求1所述的方法,其特征在于,将确定得到的各个词语中各个字分别对应的拼音,依据各个词语在文本信息中的逻辑顺序进行缓存,得到相应的拼音序列,包括:

将确定得到的各个词语中各个字分别对应的拼音,查找音节列表,得到各个拼音分别对应的第一索引值,所述音节列表包括各个带音调的拼音分别与相应第一索引值间的对应关系或者各个不带音调的拼音分别与相应第一索引值间的对应关系;

依据各个词语在文本信息中的逻辑顺序,缓存各个拼音分别对应的第一索引值,得到相应的第一索引序列。

3. 根据权利要求2所述的方法,其特征在于,对所述拼音序列进行语音合成处理,包括:

根据所述第一索引序列中的各个第一索引值,查找所述音节列表,得到所述第一索引序列对应的拼音序列;

基于所述拼音序列进行语音合成。

4. 根据权利要求1所述的方法,其特征在于,对所述文本信息进行分词,得到至少一个词语,包括:

对所述文本信息进行分词,得到至少一个词语及所述至少一个词语的词性;

在得到所述至少一个词语的词性之后,该方法还包括:

查找词性列表,得到各个词语的词性分别对应的第二索引值,并基于各个词语在文本信息中的逻辑顺序缓存各个第二索引值,得到相应的第二索引序列,所述词性列表包括各个词性分别与相应第二索引值间的对应关系;

对所述拼音序列进行语音合成处理,包括:

基于第二索引序列,对所述拼音序列进行语音合成处理。

5. 根据权利要求4所述的方法,其特征在于,基于第二索引序列,对所述拼音序列进行语音合成处理,包括:

根据所述第二索引序列中的各个第二索引值,查找所述词性列表,得到各个第二索引值分别对应的词性;

根据拼音序列及拼音序列对应的各个词语的词性,确定拼音序列中包括的各个拼音间的停顿级别;

根据所述停顿级别,对所述拼音序列进行语音合成处理。

6. 一种语音合成装置,其特征在于,包括:

获取模块,用于获取待合成语音的文本信息;

处理模块,包括分词子模块、第一确定子模块、第二确定子模块与语音合成子模块;所述分词子模块,用于对所述文本信息进行分词,得到至少一个词语;所述第一确定子模块,用于基于各个词语在文本信息中的逻辑顺序,根据利用比特表示节点与节点之间的连接关系的字典树,依次确定各个词语中各个字分别对应的拼音,所述字典树的节点包括字的节点以及词语的节点,词语的节点中携带相应词语中各个字分别对应的拼音;所述第二确定子模块,用于将确定得到的各个词语中各个字分别对应的拼音,依据各个词语在文本信息中的逻辑顺序进行缓存,得到相应的拼音序列;所述语音合成子模块,用于对所述拼音序列进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树;

输出模块,用于输出所述语音信号。

7. 一种电子设备,其特征在于,包括:

至少一个处理器;

以及与所述处理器连接的至少一个存储器、总线;其中,

所述处理器、存储器通过所述总线完成相互间的通信;

所述处理器用于调用所述存储器中的程序指令,以执行权利要求1至权利要求5中任一项所述的语音合成方法。

8. 一种非暂态计算机可读存储介质,其特征在于,所述非暂态计算机可读存储介质存储计算机指令,所述计算机指令使所述计算机执行权利要求1至权利要求5中任一项所述的语音合成方法。

语音合成方法、装置、电子设备及非暂态计算机存储介质

技术领域

[0001] 本发明实施例涉及语音处理技术领域,特别是涉及一种语音合成方法、装置、电子设备及非暂态计算机存储介质。

背景技术

[0002] 语音合成技术又称文语转换(TTS,Text-to-Speech)技术,即将文字转换为语音的技术,涉及声学、语言学、数字信号处理技术、多媒体技术等多个学科技术,是中文信息处理领域的一项前沿技术,该技术赋予智能设备(例如智能音响、机器人等)像人一样自如说话的能力,使用户与智能设备之间的信息沟通更加舒服自然。

[0003] 随着移动互联网和人工智能技术的快速发展,很多交互场景中需要智能设备直接或间接的与用户进行语音的智能交互,例如电子导航中的语音提示,机器人答题环节中的抢答和语音答复等,另外,语音播报、听小说、听新闻等一系列语音合成的场景越来越多。目前,语音合成系统在对文本信息进行语音合成时,首先对输入的文本进行归一化预处理,然后对文本进行分词、词性标注、注音等操作,再对文本进行韵律等级的预测,以及预测声学参数,最后输出最终的语音结果。

[0004] 发明人在具体实施过程中,发现现有技术中存在如下缺陷:在基于传统字典树对文本分词后的词语进行词性标注、注音、语音合成等环节中,需要存储字典树,但是由于传统字典树利用指针表示节点与节点之间的连接关系,因而占用了极大的内存空间,降低了整个内存空间的使用效率。

发明内容

[0005] 有鉴于此,本发明实施例提供了一种语音合成方法、装置、电子设备及非暂态计算机存储介质,能够极大减小占用的内存空间,提高内存空间的使用效率。

[0006] 为了解决上述问题,本发明实施例主要提供如下技术方案:

[0007] 第一方面,本发明实施例提供了一种语音合成方法,该方法包括:

[0008] 获取待合成语音的文本信息;

[0009] 基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树;

[0010] 输出语音信号。

[0011] 第二方面,本发明实施例还提供了一种语音合成装置,该装置包括:

[0012] 获取模块,用于获取待合成语音的文本信息;

[0013] 处理模块,用于基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树;

[0014] 输出模块,用于输出语音信号。

[0015] 第三方面,本发明实施例还提供了一种电子设备,包括:

- [0016] 至少一个处理器；
- [0017] 以及与处理器连接的至少一个存储器、总线；其中，
- [0018] 处理器、存储器通过总线完成相互间的通信；
- [0019] 处理器用于调用存储器中的程序指令，以执行上述的语音合成方法。
- [0020] 第四方面，本发明实施例还提供了一种非暂态计算机可读存储介质，其中，非暂态计算机可读存储介质存储计算机指令，计算机指令使计算机执行上述的语音合成方法。
- [0021] 借由上述技术方案，本发明实施例提供的技术方案至少具有下列优点：
- [0022] 本发明实施例提供的语音合成方法，获取待合成语音的文本信息，为后续对文本信息进行语音合成处理奠定基础；基于预先建立的语音合成模型，对文本信息进行语音合成处理，得到相应的语音信号，语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树，从而利用比特表示节点与节点之间的连接关系的字典树替代传统的利用指针表示节点与节点之间的连接关系的字典树，由于指针占用的字节数远大于比特，因而利用比特表示节点与节点之间的连接关系，极大减小了占用的内存空间，利于后续提高内存空间的使用效率；输出语音信号，从而将文本信息转换为语音信号。
- [0023] 上述说明仅是本发明实施例技术方案的概述，为了能够更清楚了解本发明实施例的技术手段，而可依照说明书的内容予以实施，并且为了让本发明实施例的上述和其它目的、特征和优点能够更明显易懂，以下特举本发明实施例的具体实施方式。

附图说明

- [0024] 通过阅读下文优选实施方式的详细描述，各种其他的优点和益处对于本领域普通技术人员将变得清楚明了。附图仅用于示出优选实施方式的目的，而并不认为是对本发明实施例的限制。而且在整个附图中，用相同的参考符号表示相同的部件。在附图中：
- [0025] 图1示出了本发明实施例提供的语音合成方法的流程示意图；
- [0026] 图2示出了本发明实施例的利用比特表示节点与节点之间的连接关系的字典树；
- [0027] 图3示出了本发明实施例的根据字根查找字典树的示意图；
- [0028] 图4示出了本发明实施例提供的语音合成装置的基本结构示意图；
- [0029] 图5示出了本发明实施例提供的语音合成装置的详细结构示意图；
- [0030] 图6示出了本发明实施例提供的电子设备的结构示意图。

具体实施方式

[0031] 下面将参照附图更详细地描述本公开的示例性实施例。虽然附图中显示了本公开的示例性实施例，然而应当理解，可以以各种形式实现本公开而不应被这里阐述的实施例所限制。相反，提供这些实施例是为了能够更透彻地理解本公开，并且能够将本公开的范围完整的传达给本领域的技术人员。

[0032] 实施例一

[0033] 本发明实施例提供了一种语音合成方法，如图1所示，包括：步骤S110，获取待合成语音的文本信息；步骤S120，基于预先建立的语音合成模型，对文本信息进行语音合成处理，得到相应的语音信号，语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树；步骤S130：输出语音信号。

[0034] 本发明的实施例可以应用于数据处理设备,这里的数据处理设备可以是具有数据处理功能的智能设备,并可以直接或间接的与用户进行交互,例如可以是移动终端、机器人等,也可以是后台的服务器等,当数据处理设备为服务器这种不具有直接与用户进行交互的设备时,与用户进行交互的设备可以与该服务器数据连接,并通过与服务器的数据交互实现与用户的交互。

[0035] 下面以数据处理设备是具有数据处理功能的智能设备为例,对本发明实施例的语音合成方法进行详细介绍,如下所示:

[0036] 步骤S110,获取待合成语音的文本信息。

[0037] 具体地,智能设备可以获取预先存储的文本信息,将该获取到的文本信息作为待合成语音的文本信息,比如在人机交互场景中,根据用户当前的语音信息(例如“早上好”),从预先存储的文本信息中获取合适的文本信息,例如获取到“早上好,祝你今天有个好心情”这一文本信息,并将该文本信息合成语音,以与用户进行沟通、交流。

[0038] 智能设备也可以通过接收用户输入的待合成语音的文本信息的方式,来获取待合成语音的文本信息,比如用户将好友发来的短信等文本信息输入到智能设备,又比如用户将小说的文本信息输入到智能设备,即智能设备获取用户输入的短信文本、小说文本等。

[0039] 步骤S120,基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树。

[0040] 具体地,智能设备基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,例如智能设备基于预先建立的语音合成模型,对获取到的“早上好,祝你今天有个好心情”这一文本信息,进行语音合成处理,得到相应的语音信号,又例如智能设备基于预先建立的语音合成模型,对好友发来的短信“今天有暴雨,注意出行安全”这一文本信息进行语音合成处理,得到相应的语音信号,再例如智能设备基于预先建立的语音合成模型,对用户输入的小说“平凡的世界……”这一文本信息进行语音合成处理,得到相应的语音信号。

[0041] 进一步地,上述的语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树,比如可以用一个或几个比特(例如“0”或“1”)表示节点与节点之间的连接关系,如图2所示,“01”表示节点“中”与节点“闻”之间的连接关系,“10”表示节点“中”与节点“文”之间的连接关系,“11”表示节点“中”与节点“医”之间的连接关系,代替传统的利用指针表示节点与节点之间的连接关系的字典树,由于一个指针占用8个字节,即64比特,而整个字典树需要用若干个指针来表示节点与节点之间的连接关系,因而存储该若干个指针需要占用极大的内存空间,远远大于存储比特占用的内存空间,从而利用比特表示节点与节点之间的连接关系,极大减小了占用的内存空间,利于后续提高内存空间的使用效率。

[0042] 步骤S130:输出语音信号。

[0043] 具体地,在基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号之后,智能设备输出语音信号,例如播放该语音信号。

[0044] 本发明实施例提供的语音合成方法,与现有技术相比,获取待合成语音的文本信息,为后续对文本信息进行语音合成处理奠定基础;基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树,从而利用比特表示节点与节点之间的连接关系的字典树替代

传统的利用指针表示节点与节点之间的连接关系的字典树,由于指针占用的字节数远大于比特,因而利用比特表示节点与节点之间的连接关系,极大减小了占用的内存空间,利于后续提高内存空间的使用效率;输出语音信号,从而将文本信息转换为语音信号。

[0045] 实施例二

[0046] 本发明实施例提供了另一种可能的实现方式,在实施例一的基础上,还包括实施例二所示的方法,其中,

[0047] 步骤S120包括步骤S1201(图中未标注)、步骤S1202(图中未标注)、步骤S1203(图中未标注)及步骤S1204(图中未标注),其中,

[0048] 步骤S1201:对文本信息进行分词,得到至少一个词语。

[0049] 步骤S1202:基于各个词语在文本信息中的逻辑顺序,根据利用比特表示节点与节点之间的连接关系的字典树,依次确定各个词语中各个字分别对应的拼音,字典树的节点包括字的节点以及词语的节点,词语的节点中携带相应词语中各个字分别对应的拼音。

[0050] 步骤S1203:将确定得到的各个词语中各个字分别对应的拼音,依据各个词语在文本信息中的逻辑顺序进行缓存,得到相应的拼音序列。

[0051] 步骤S1204:对拼音序列进行语音合成处理。

[0052] 具体地,在基于预先建立的语音合成模型,对文本信息进行语音合成处理的过程中,首先对待合成语音的文本信息进行分词,得到至少一个词语,例如对待合成语音的文本信息“狂风暴雨”进行分词,得到“狂风”与“暴雨”两个词语,接着基于“狂风”与“暴雨”这两个词语在文本信息中的逻辑顺序,“狂风”与“暴雨”紧挨着且“狂风”在“暴雨”之前,根据利用比特表示节点与节点之间的连接关系的字典树,依次确定“狂风”与“暴雨”中各个字分别对应的拼音,其中,字典树的节点包括字的节点以及词语的节点,词语的节点中携带相应词语中各个字分别对应的拼音,如图3所示。

[0053] 进一步地,图3简要示出了字典树的一个子分支,在图3中,字典树的根节点(即字根)是“狂”字,与根节点相连接的节点为各个字的节点,例如“风”字节点、“妄”字节点及“奔”字节点等,其中,“狂”与“风”构成“狂风”这一词语,且该词语与文本信息中的“狂风”相匹配,即通过查找字典树得到待合成语音的文本信息中的词语,其中,字典树中包括“狂风”这一词语的节点,同时在相应的词语节点中携带“狂风”这一词语的带音调或不带单调的拼音“kuangfeng”,接着字典树中“狂风”这一词语节点与“暴”这一字节点相连接,构成“狂风暴”这一词语,字典树中包括“狂风暴”这一词语的节点,同时在该词语节点中携带“狂风暴”这一词语的带音调或不带单调的拼音“kuangfengbao”,再接着字典树中“狂风暴”这一词语节点与“雨”这一字节点相连接,构成“狂风暴雨”这一词语,字典树中包括“狂风暴雨”这一词语的节点,同时在该词语节点中携带“狂风暴雨”这一词语的带音调或不带单调的拼音“kuangfengbaoyu”。至此,通过查找字典树得到待合成语音的文本信息“狂风暴雨”,同时根据字典树中各个词语字节中携带的带音调或不带单调的词语拼音,得到对文本信息分词后的各个词语中各个字分别对应的拼音“kuang”、“feng”、“bao”与“yu”。

[0054] 进一步地,将确定得到的文本信息的各个词语“狂风”与“暴雨”中各个字分别对应的拼音“kuang”、“feng”、“bao”与“yu”,依据各个词语在文本信息中的逻辑顺序(即“狂风”与“暴雨”紧挨着且“狂风”在“暴雨”之前)进行缓存,直到得到完整的拼音序列“kuangfengbaoyu”,随后对得到的完整的拼音序列“kuangfengbaoyu”进行语音合成处理,

即可得到相应的语音信号。

[0055] 对于本发明实施例,通过利用比特表示节点与节点之间的连接关系的字典树代替传统的利用指针表示节点与节点之间的连接关系的字典树,极大减小了字典树占用的内存空间,利于后续提高内存空间的使用效率。

[0056] 实施例三

[0057] 本发明实施例提供了另一种可能的实现方式,在实施例二的基础上,还包括实施例三所示的方法。

[0058] 步骤S1203具体包括步骤S12031(图中未标注)与步骤S12032(图中未标注),其中,

[0059] 步骤S12031:将确定得到的各个词语中各个字分别对应的拼音,查找音节列表,得到各个拼音分别对应的第一索引值,音节列表包括各个带音调的拼音分别与相应第一索引值间的对应关系或者各个不带音调的拼音分别与相应第一索引值间的对应关系。

[0060] 步骤S12032:依据各个词语在文本信息中的逻辑顺序,缓存各个拼音分别对应的第一索引值,得到相应的第一索引序列。

[0061] 步骤S1204具体包括步骤S12041(图中未标注)与步骤S12042(图中未标注),其中,

[0062] 步骤S12041:根据第一索引序列中的各个第一索引值,查找音节列表,得到第一索引序列对应的拼音序列。

[0063] 步骤S12042:基于拼音序列进行语音合成。

[0064] 具体地,依次缓存文本信息的各个词语中各个字分别对应的拼音,将会占用大量的内存空间,一方面存在相同字的拼音重复缓存的情况,另一方面缓存一个字的拼音需要占用较多的字节,例如“kuang”这一拼音占用6上字节,尤其当文本信息较长时,此时可能需要缓存数以千计、数以万计的拼音,将极大占用内存空间。

[0065] 进一步地,为了减小缓存拼音序列时占用的内存空间,本发明实施例用音节列表的方式缓存所有的拼音,该拼音可以是带音调的拼音也可以是不带音调的拼音,其中,带音调的拼音总共大约有1400个,不带音调的拼音总共大约有400个,换言之,可以在音节列表中将所有带音调的拼音做一个空间的映射,映射为一系列的索引值,如表1所示:

[0066] 表1音节列表

[0067]

bao	ben	bu	feng	kuang	pao	yu
1	2	3	4	5	6	7

[0068] 进一步地,将确定得到的各个词语中各个字分别对应的拼音,例如“kuang”、“feng”、“bao”与“yu”,查找音节列表,得到各个拼音分别对应的索引值(即上述的第一索引值),其中,根据表1可以看出,拼音“bao”映射为索引值1,拼音“feng”映射为索引值4,拼音“kuang”映射为索引值5,拼音“yu”映射为索引值7,得到各个拼音分别对应的第一索引值之后,在缓存文本信息的各个词语中各个字分别对应的拼音时,可以只缓存各个拼音分别对应的索引值,例如1、4、5与7,其中,可以依据各个词语在文本信息中的逻辑顺序,缓存各个拼音分别对应的索引值,得到相应的索引序列“5417”(即上述的第一索引序列),即用“5417”这一索引序列代替拼音序列“kuangfengbaoyu”,从而极大减小占用的内存空间,便于提高内存空间的使用效率。

[0069] 进一步地,通常是在缓存完整的拼音序列后再对该完整的拼音序列进行语音合成,即需要在缓存完整的索引序列后,再基于该索引序列进行语音合成处理,其中,在基于

索引序列进行语音合成处理时,可以根据上述的音节列表查找索引序列“5417”中各个索引值分别对应的拼音,即索引值1对应拼音“bao”,索引值4对应拼音“feng”,索引值5对应拼音“kuang”,索引值7对应拼音“yu”,从而得到索引序列“5417”对应的拼音序列“kuangfengbaoyu”,随后对得到的完整的拼音序列“kuangfengbaoyu”进行语音合成处理,即可得到相应的语音信号。

[0070] 对于本发明实施例,通过缓存各个拼音的索引值来代替各个拼音,从而极大减小缓存拼音序列时占用的内存空间,进一步减少语音合成过程中占用的内存空间,实现内存的优化存储,提高内存空间的使用效率。

[0071] 实施例四

[0072] 本发明实施例提供了另一种可能的实现方式,在实施例二的基础上,还包括实施例四所示的方法。

[0073] 步骤S1201具体可以为:对文本信息进行分词,得到至少一个词语及至少一个词语的词性。

[0074] 在步骤S1201之后还包括步骤S1205(图中未标注):查找词性列表,得到各个词语的词性分别对应的第二索引值,并基于各个词语在文本信息中的逻辑顺序缓存各个第二索引值,得到相应的第二索引序列,词性列表包括各个词性分别与相应第二索引值间的对应关系。

[0075] 步骤S1204具体可以为:基于第二索引序列,对拼音序列进行语音合成处理。

[0076] 步骤S1204具体包括步骤S12043(图中未标注)、步骤S12044(图中未标注)及步骤S12045(图中未标注),其中,

[0077] 步骤S12043:根据第二索引序列中的各个第二索引值,查找词性列表,得到各个第二索引值分别对应的词性;

[0078] 步骤S12044:根据拼音序列及拼音序列对应的各个词语的词性,确定拼音序列中包括的各个拼音间的停顿级别;

[0079] 步骤S12044:根据停顿级别,对拼音序列进行语音合成处理。

[0080] 具体地,在对文本信息“狂风暴雨”进行分词时,除了得到至少一个词语“狂风”与“暴雨”外,还可以得到该至少一个词语的词性,即得到“狂风”这一词语的词性为名词,“暴雨”这一词语的词性为名词。

[0081] 进一步地,在得到词语“狂风”与“暴雨”的词性后,不仅需要缓存词语“狂风”与“暴雨”中各个字的拼音,还需要缓存词语“狂风”与“暴雨”的词性,其中,词性包括名词、动词、形容词等等,在缓存词性的过程中,也会占用较多的内存空间,一方面存在相同词性重复缓存的情况,另一方面缓存词性占用较多的字节。

[0082] 进一步地,为了减小缓存词性时占用的内存空间,本发明实施例用词性列表的方式缓存所有的词性,可以在词性列表中将所有词性做一个空间的映射,映射为一系列的索引值(即上述的第二索引值),如表2所示:

[0083] 表2词性列表

[0084]

名词	动词	形容词	副词	代词
a	b	c	d	e

[0085] 进一步地,将确定得到的各个词语分别对应的词性,例如“名词”、“动词”等,查找

词性列表,得到各个词性分别对应的索引值(即上述的第二索引值),其中,根据表2可以看出,名词映射为索引值a,动词映射为索引值b,形容词映射为索引值c等等,得到各个词语分别对应的词性后,在缓存各个词语的词性时,可以只缓存各个词性分别对应的索引值,例如a、b及c等,其中,可以依据各个词语在文本信息中的逻辑顺序,缓存各个词性分别对应的索引值,得到相应的索引序列“aa”(即上述的第二索引序列),即用“aa”这一索引序列代替词性序列“名词名词”,从而极大减小占用的内存空间,便于提高内存空间的使用效率。

[0086] 需要说明的是,上述的第二索引值可以与第一索引值相同,即第二索引值也使用数字1、2、3等表示,也可以与第一索引值不同,例如第二索引值使用字母a、b、c等表示。

[0087] 进一步地,在对拼音序列进行语音合成处理时,可以基于第二索引序列“aa”对拼音序列进行语音合成处理。其中,在基于第二索引序列“aa”对拼音序列进行语音合成处理的过程中,可以根据上述的词性列表查找索引序列“aa”中各个索引值分别对应的词性,即第一个索引值a对应名词,第二个索引值b对应名词。

[0088] 进一步地,得到索引序列“aa”中的各个索引值分别对应的词性后,可以根据上述得到的拼音序列及拼音序列对应的各个词语的词性,确定拼音序列中包括的各个拼音间的停顿级别,其中,停顿级别可以描述为停顿时长,例如“kuang”与“feng”间停顿5毫秒,“bao”与“yu”间停顿5毫秒,“kuangfeng”与“baoyu”间停顿10毫秒。

[0089] 进一步地,在确定拼音序列中包括的各个拼音间的停顿级别后,可以根据停顿级别,对得到的完整的拼音序列进行语音合成处理,从而使得合成的语音能够更好地模拟人类语言的语音,使得合成的语音能够高低起伏、停顿转折,十分和谐而有节奏。

[0090] 对于本发明实施例,通过缓存各个词性的索引值来代替各个词性,从而极大减小缓存词性序列时占用的内存空间,进一步减少语音合成过程中占用的内存空间,实现内存的优化存储,提高内存空间的使用效率。

[0091] 实施例五

[0092] 图4为本发明实施例提供的一种语音合成装置的结构示意图,如图4所示,该装置40可以包括获取模块41、处理模块42及输出模块43,其中,

[0093] 获取模块41用于获取待合成语音的文本信息;

[0094] 处理模块42用于基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树;

[0095] 输出模块43用于输出语音信号。

[0096] 具体地,处理模块42包括分词子模块421、第一确定子模块422、第二确定子模块423与语音合成子模块424,如图5所示,其中,

[0097] 分词子模块421用于对文本信息进行分词,得到至少一个词语;

[0098] 第一确定子模块422用于基于各个词语在文本信息中的逻辑顺序,根据利用比特表示节点与节点之间的连接关系的字典树,依次确定各个词语中各个字分别对应的拼音,字典树的节点包括字的节点以及词语的节点,词语的节点中携带相应词语中各个字分别对应的拼音;

[0099] 第二确定子模块423用于将确定得到的各个词语中各个字分别对应的拼音,依据各个词语在文本信息中的逻辑顺序进行缓存,得到相应的拼音序列;

[0100] 语音合成子模块424用于对拼音序列进行语音合成处理。

[0101] 进一步地,第二确定子模块423包括第一查找子单元4231(图中未标注)与缓存子单元4232(图中未标注),其中,

[0102] 第一查找子单元4231用于将确定得到的各个词语中各个字分别对应的拼音,查找音节列表,得到各个拼音分别对应的第一索引值,音节列表包括各个带音调的拼音分别与相应第一索引值间的对应关系或者各个不带音调的拼音分别与相应第一索引值间的对应关系;

[0103] 缓存子单元4232用于依据各个词语在文本信息中的逻辑顺序,缓存各个拼音分别对应的第一索引值,得到相应的第一索引序列。

[0104] 进一步地,语音合成子模块424包括第二查找子单元4241(图中未标注)与第一合成子单元4242(图中未标注),其中,

[0105] 第二查找子单元4241用于根据第一索引序列中的各个第一索引值,查找音节列表,得到第一索引序列对应的拼音序列;

[0106] 第一合成子单元4242用于基于拼音序列进行语音合成。

[0107] 进一步地,分词子模块421具体用于对文本信息进行分词,得到至少一个词语及至少一个词语的词性;

[0108] 处理模块还包括词性确定子模块425,如图5所示,其中,词性确定子模块425用于查找词性列表,得到各个词语的词性分别对应的第二索引值,并基于各个词语在文本信息中的逻辑顺序缓存各个第二索引值,得到相应的第二索引序列,词性列表包括各个词性分别与相应第二索引值间的对应关系;

[0109] 语音合成子模块424具体用于基于第二索引序列,对拼音序列进行语音合成处理。

[0110] 进一步地,语音合成子模块424包括第三查找子单元4241(图中未标注)、停顿级别确定子单元4242(图中未标注)与第二合成子单元4243(图中未标注),其中,

[0111] 第三查找子单元4241用于根据第二索引序列中的各个第二索引值,查找词性列表,得到各个第二索引值分别对应的词性;

[0112] 停顿级别确定子单元4242用于根据拼音序列及拼音序列对应的各个词语的词性,确定拼音序列中包括的各个拼音间的停顿级别;

[0113] 第二合成子单元4243用于根据停顿级别,对拼音序列进行语音合成处理。

[0114] 本发明实施例提供的语音合成装置,与现有技术相比,获取待合成语音的文本信息,为后续对文本信息进行语音合成处理奠定基础;基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树,从而利用比特表示节点与节点之间的连接关系的字典树替代传统的利用指针表示节点与节点之间的连接关系的字典树,由于指针占用的字节数远大于比特,因而利用比特表示节点与节点之间的连接关系,极大减小了占用的内存空间,利于后续提高内存空间的使用效率;输出语音信号,从而将文本信息转换为语音信号。

[0115] 由于本发明实施例所介绍的语音合成装置为可以执行本发明实施例中的语音合成方法的装置,故而基于本发明实施例中所介绍的语音合成方法,本领域所属技术人员能够了解本实施例的语音合成装置的具体实施方式以及其各种变化形式,所以在此对于该语音合成装置如何实现本发明实施例中的语音合成方法不再详细介绍。只要本领域所属技术

人员实施本发明实施例中语音合成方法所采用的装置,都属于本发明所欲保护的范围。

[0116] 实施例六

[0117] 本发明实施例提供了一种电子设备,如图6所示,图6所示的电子设备60包括:处理器61和存储器62。其中,处理器61和存储器62相连,如通过总线63相连。进一步地,电子设备60还可以包括收发器64(图中未标注)。需要说明的是,实际应用中收发器64不限于一个,该电子设备60的结构并不构成对本发明实施例的限定。

[0118] 其中,处理器61应用于本发明实施例中,用于实现图4或图5所示的获取模块、处理模块及输出模块的功能。

[0119] 处理器61可以是CPU,通用处理器,DSP,ASIC,FPGA或者其他可编程逻辑器件、晶体管逻辑器件、硬件部件或者其任意组合。其可以实现或执行结合本发明公开内容所描述的各种示例性的逻辑方框,模块和电路。处理器61也可以是实现计算功能的组合,例如包含一个或多个微处理器组合,DSP和微处理器的组合等。

[0120] 总线63可包括一通路,在上述组件之间传送信息。总线63可以是PCI总线或EISA总线等。总线63可以分为地址总线、数据总线、控制总线等。为便于表示,图6中仅用一条粗线表示,但并不表示仅有一根总线或一种类型的总线。

[0121] 存储器62可以是ROM或可存储静态信息和指令的其他类型的静态存储设备,RAM或者可存储信息和指令的其他类型的动态存储设备,也可以是EEPROM、CD-ROM或其他光盘存储、光碟存储(包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质,但不限于此。

[0122] 存储器62用于存储执行本发明方案的应用程序代码,并由处理器61来控制执行。处理器61用于执行存储器62中存储的应用程序代码,以实现图4或图5所示实施例提供的语音合成装置的动作。

[0123] 本发明实施例提供的电子设备,包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序,处理器执行程序时,与现有技术相比,可实现:获取待合成语音的文本信息,为后续对文本信息进行语音合成处理奠定基础;基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树,从而利用比特表示节点与节点之间的连接关系的字典树替代传统的利用指针表示节点与节点之间的连接关系的字典树,由于指针占用的字节数远大于比特,因而利用比特表示节点与节点之间的连接关系,极大减小了占用的内存空间,利于后续提高内存空间的使用效率;输出语音信号,从而将文本信息转换为语音信号。

[0124] 本实施例提供一种非暂态计算机可读存储介质,所述非暂态计算机可读存储介质存储计算机指令,所述计算机指令使所述计算机执行上述各方法实施例所提供的方法。与现有技术相比,获取待合成语音的文本信息,为后续对文本信息进行语音合成处理奠定基础;基于预先建立的语音合成模型,对文本信息进行语音合成处理,得到相应的语音信号,语音合成模型包括利用比特表示节点与节点之间的连接关系的字典树,从而利用比特表示节点与节点之间的连接关系的字典树替代传统的利用指针表示节点与节点之间的连接关系的字典树,由于指针占用的字节数远大于比特,因而利用比特表示节点与节点之间的连接关系,极大减小了占用的内存空间,利于后续提高内存空间的使用效率;输出语音信号,

从而将文本信息转换为语音信号。

[0125] 本领域内的技术人员应明白,本发明的实施例可提供为方法、系统、或计算机程序产品。因此,本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0126] 本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0127] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0128] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0129] 在一个典型的配置中,计算设备包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0130] 存储器可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。存储器是计算机可读介质的示例。

[0131] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁磁盘存储或其他磁性存储设备或任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0132] 还需要说明的是,术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含,从而使得包括一系列要素的过程、方法、商品或者设备不仅包括那些要素,而且还包括没有明确列出的其他要素,或者是还包括为这种过程、方法、商品或者设备所固有的要素。在没有更多限制的情况下,由语句“包括一个……”限定的要素,并不排除在包括要素的过程、方法、商品或者设备中还存在另外的相同要素。

[0133] 本领域技术人员应明白,本发明的实施例可提供为方法、系统或计算机程序产品。

因此,本发明可采用完全硬件实施例、完全软件实施例或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0134] 以上仅为本发明的实施例而已,并不用于限制本发明。对于本领域技术人员来说,本发明可以有各种更改和变化。凡在本发明的精神和原理之内所作的任何修改、等同替换、改进等,均应包含在本发明的权利要求范围之内。

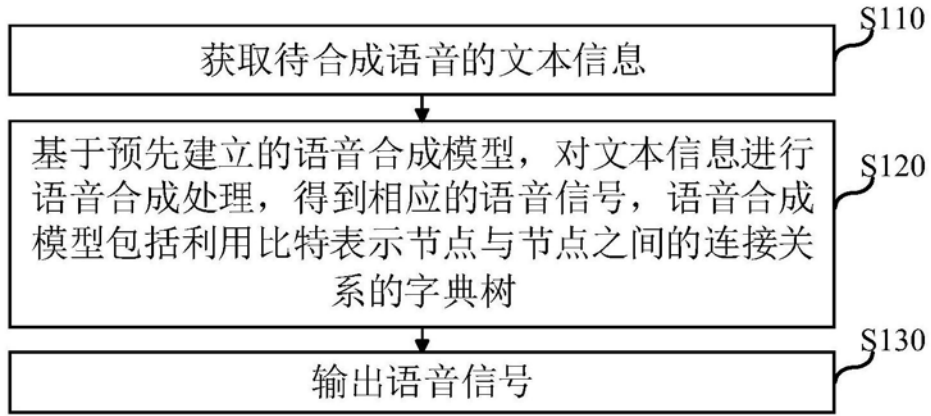


图1

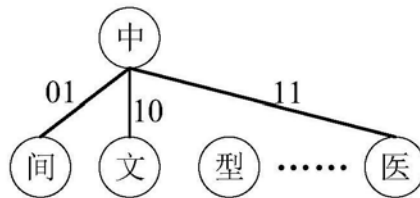


图2

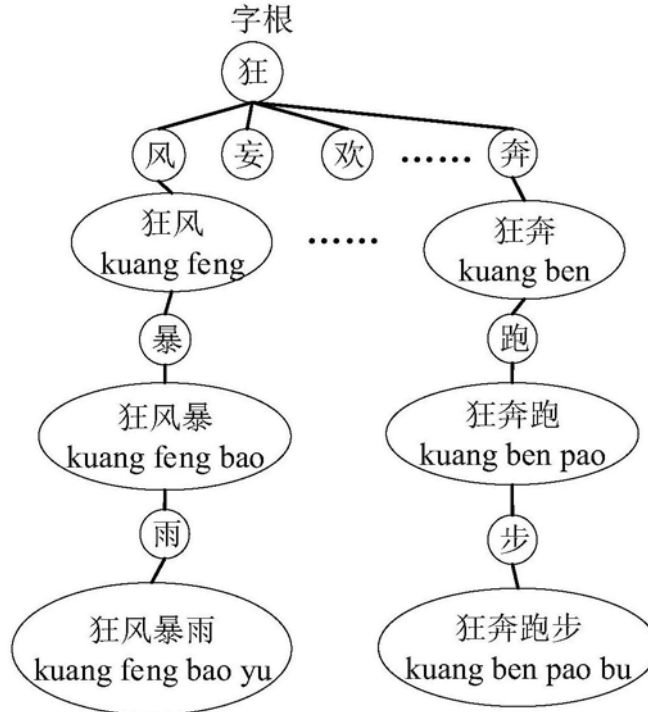


图3

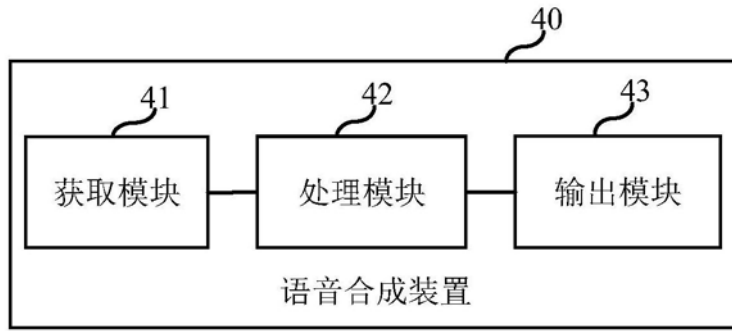


图4

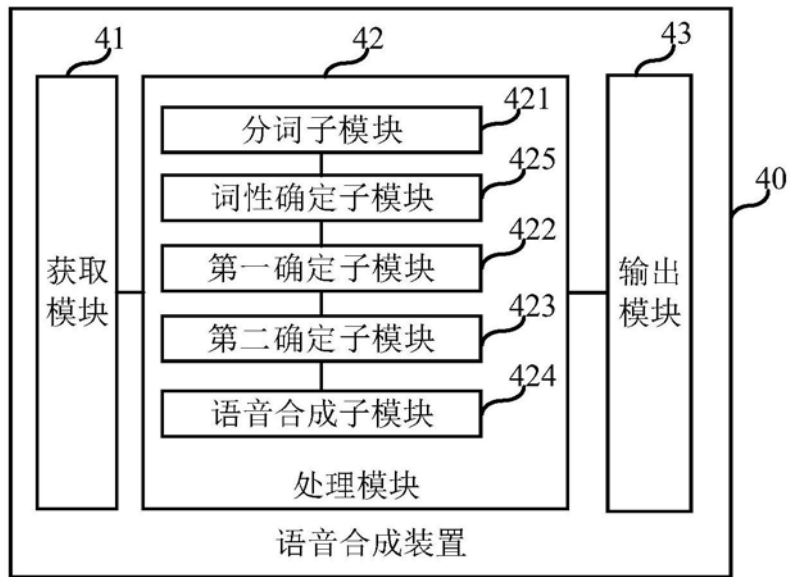


图5

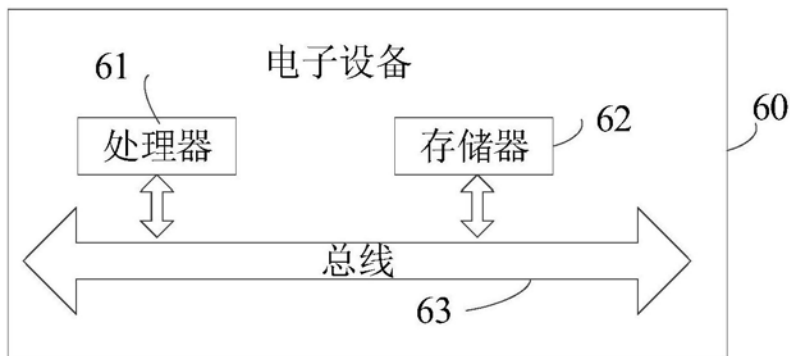


图6