



(12)发明专利申请

(10)申请公布号 CN 106504742 A

(43)申请公布日 2017. 03. 15

(21)申请号 201610999015.2

(22)申请日 2016.11.14

(71)申请人 海信集团有限公司

地址 266071 山东省青岛市市南区东海西路17号

(72)发明人 匡涛 任晓楠 王峰 张大钊

(74)专利代理机构 北京律智知识产权代理有限公司 11438

代理人 邢雪红 乔彬

(51) Int. Cl.

G10L 13/02(2013.01)

H04L 29/06(2006.01)

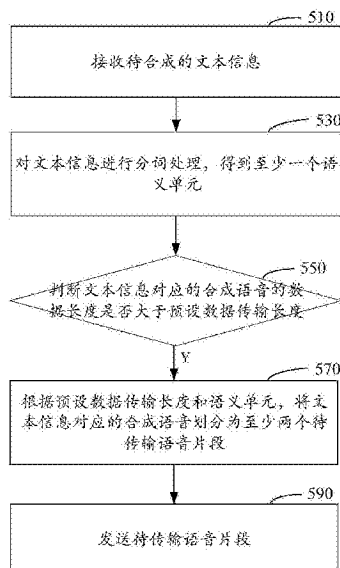
权利要求书3页 说明书15页 附图11页

(54)发明名称

合成语音的传输方法、云端服务器和终端设备

(57)摘要

本公开涉及了一种合成语音的传输方法、云端服务器和终端设备。所述合成语音的传输方法,包括:接收待合成的文本信息;对所述文本信息进行分词处理,得到至少一个语义单元;判断所述文本信息对应的合成语音的数据长度是否大于预设数据传输长度;若为是,则根据所述预设数据传输长度和语义单元,将所述文本信息对应的合成语音划分为至少两个待传输语音片段,所述待传输语音片段是若干个语义单元对应的合成语音;发送所述待传输语音片段。由于待传输语音片段是由若干个语义单元对应的合成语音组成的,因此,无论网络环境是否异常,该待传输的语音段都将保持文本信息原有的语义结构,从而保证了经传输的合成语音的可理解性。



1. 一种应用于云端服务器的合成语音的传输方法,其特征在于,包括:
接收待合成的文本信息;
对所述文本信息进行分词处理,得到至少一个语义单元;
判断所述文本信息对应的合成语音的数据长度是否大于预设数据传输长度;
若为是,则根据所述预设数据传输长度和语义单元,将所述文本信息对应的合成语音划分为至少两个待传输语音片段,所述待传输语音片段是若干个语义单元对应的合成语音;
发送所述待传输语音片段。
2. 如权利要求1所述的方法,其特征在于,所述根据所述预设数据传输长度和语义单元,将所述文本信息对应的合成语音划分为至少两个待传输语音片段的步骤包括:
判断所述文本信息中第一个语义单元对应的合成语音的数据长度是否大于所述预设数据传输长度;
若为否,则将所述第一个语义单元和第二个语义单元分别对应的合成语音的数据长度累加,得到第一数据长度累加和;
进一步判断所述第一数据长度累加和是否大于所述预设数据传输长度;
若为是,则以所述第一个语义单元对应的合成语音作为所述待传输语音片段。
3. 如权利要求1所述的方法,其特征在于,所述根据所述预设数据传输长度和语义单元,将所述文本信息对应的合成语音划分为至少两个待传输语音片段的步骤包括:
将所述文本信息对应的合成语音的数据长度减去倒数第一个语义单元对应的合成语音的数据长度,得到第一数据长度差值;
判断所述第一数据长度差值是否大于所述预设数据传输长度;
若为否,则将所述倒数第一个语义单元之前的所有语义单元对应的合成语音作为所述待传输语音片段。
4. 如权利要求1所述的方法,其特征在于,所述判断所述文本信息对应的合成语音的数据长度是否大于预设数据传输长度的步骤之前,所述方法还包括:
根据汉语发音时长计算所述文本信息包括的各语义单元的发音时长;
根据所述文本信息包括的各语义单元的发音时长之和,得到所述文本信息的发音时长;
根据所述文本信息的发音时长,确定所述文本信息对应的合成语音的数据长度。
5. 一种应用于云端服务器的合成语音的传输方法,其特征在于,包括:
接收待合成的文本信息;
对所述文本信息进行分词处理,得到至少一个语义单元;
根据预设数据传输长度和所述语义单元生成待传输语音片段,所述待传输语音片段是若干个语义单元对应的合成语音,且所述若干个语义单元对应的合成语音的数据长度之和不大于所述预设数据传输长度;
发送所述待传输语音片段。
6. 如权利要求5所述的方法,其特征在于,所述根据所述预设数据传输长度和语义单元生成待传输语音片段的步骤包括:
判断所述文本信息中第一个语义单元对应的合成语音的数据长度是否大于所述预设

数据传输长度；

若为否，则将所述第一个语义单元和第二个语义单元分别对应的合成语音的数据长度累加，得到第一数据长度累加和；

进一步判断所述第一数据长度累加和是否大于所述预设数据传输长度；

若为是，则以所述第一个语义单元对应的合成语音作为所述待传输语音片段。

7. 如权利要求6所述的方法，其特征在于，所述判断所述文本信息中第一个语义单元对应的合成语音的数据长度是否大于所述预设数据传输长度的步骤之前，所述方法还包括：

根据汉语发音时长计算所述文本信息中第一个语义单元的发音时长；

根据所述第一个语义单元的发音时长确定所述第一个语义单元对应的合成语音的数据长度。

8. 一种应用于终端设备的合成语音的传输方法，其特征在于，包括：

向云端服务器发送语音合成请求，所述语音合成请求由待合成的文本信息生成，以使所述云端服务器通过响应所述语音合成请求对所述文本信息进行语音合成；

接收所述云端服务器返回的传输语音片段，其中，所述传输语音片段是若干个语义单元对应的合成语音，且所述若干个语义单元对应的合成语音的数据长度之和不大于预设数据传输长度；

播报所述传输语音片段。

9. 一种云端服务器，其特征在于，所述云端服务器包括：

信息接收模块，用于接收待合成的文本信息；

分词处理模块，用于对所述文本信息进行分词处理，得到至少一个语义单元；

判断模块，用于判断所述文本信息对应的合成语音的数据长度是否大于预设数据传输长度；若为是，则通知语音片段划分模块；

所述语音片段划分模块，用于根据所述预设数据传输长度和语义单元，将所述文本信息对应的合成语音划分为至少两个待传输语音片段，所述待传输语音片段是若干个语义单元对应的合成语音；

发送模块，用于发送所述待传输语音片段。

10. 一种云端服务器，其特征在于，所述云端服务器包括：

信息接收模块，用于接收待合成的文本信息；

分词处理模块，用于对所述文本信息进行分词处理，得到至少一个语义单元；

语音片段生成模块，用于根据预设数据传输长度和所述语义单元生成待传输语音片段，所述待传输语音片段是若干个语义单元对应的合成语音，且所述若干个语义单元对应的合成语音的数据长度之和不大于所述预设数据传输长度；

发送模块，用于发送所述待传输语音片段。

11. 一种终端设备，其特征在于，所述终端设备包括：

发送模块，用于向云端服务器发送语音合成请求，所述语音合成请求由待合成的文本信息生成，以使所述云端服务器通过响应所述语音合成请求对所述文本信息进行语音合成；

接收模块，用于接收所述云端服务器返回的传输语音片段，其中，所述传输语音片段是若干个语义单元对应的合成语音，且所述若干个语义单元对应的合成语音的数据长度之和

不大于预设数据传输长度；

语音播报模块,用于播报所述传输语音片段。

合成语音的传输方法、云端服务器和终端设备

技术领域

[0001] 本公开涉及语音合成技术领域,尤其涉及一种合成语音的传输方法及装置。

背景技术

[0002] 语音合成技术(又称文语转换技术)是将计算机内部生成或者外部输入的文字信息转换为用户可以理解的声音信息的汉语输出技术。

[0003] 由于云端处理具有运行资源占用小等优势,因此,基于云端处理的语音合成已得到较为广泛地应用。该基于云端处理的语音合成过程包括:终端设备将待合成的文本信息发送至云端服务器,由云端服务器将该待合成的文本信息通过语音合成技术合成为合成语音,再借助于网络将合成语音返回至终端设备,以通过终端设备对接收到的合成语音进行播报,进而使得用户掌握播报内容。

[0004] 如果云端服务器待语音合成完毕之后,才一次性地将合成完毕的合成语音返回终端设备,则终端设备不仅需要等待语音合成完毕,还需要等待合成语音传输完毕,才能开始播报接收到的合成语音,因此,尚存在语音合成过程耗时过长的问题。如果将合成语音先压缩再传输,虽然缩短了合成语音的传输时长,由于终端设备还需要对接收到的合成语音解压缩后才能进行播报,而压缩与解压缩同样会消耗大量的时间,仍然无法解决语音合成过程耗时过长的问题。

[0005] 为了解决语音合成过程耗时过长的问题,利用未经编码的原始音频数据传输合成语音的PCM数据传输方法应运而生,该PCM数据传输方法能够采用固定数据传输长度对合成语音进行传输,即将合成语音划分为固定长度的若干待传输语音片段进行传输,使得云端服务器一边进行语音合成一边进行待传输语音片段的传输,而终端设备无需等待语音合成完毕,也无需等待合成语音传输完毕,仅在接收到固定长度的待传输语音片段之后即可开始播报,由此有效地缩短了语音合成过程的时长。

[0006] 然而,受限于终端设备所在的网络环境,在网络环境异常时,例如,网速(即单位时间内网络的上行/下行数据量)较差,将造成终端设备接收到的固定长度的若干待传输语音片段之间不连续,即存在随机停顿,而可能破坏了待合成的文本信息原有的语义结构,进而导致用户无法理解终端设备所播报的合成语音。

发明内容

[0007] 基于此,本公开提供一种合成语音的传输方法、云端服务器和终端设备,用于解决现有技术中在网络环境异常时经传输的合成语音的可理解性较差的问题。

[0008] 一方面,本公开提供一种应用于云端服务器的合成语音的传输方法,包括:接收待合成的文本信息;对所述文本信息进行分词处理,得到至少一个语义单元;判断所述文本信息对应的合成语音的数据长度是否大于预设数据传输长度;若为是,则根据所述预设数据传输长度和语义单元,将所述文本信息对应的合成语音划分为至少两个待传输语音片段,所述待传输语音片段是若干个语义单元对应的合成语音;发送所述待传输语音片段。

[0009] 另一方面,本公开提供一种应用于云端服务器的合成语音的传输方法,包括:接收待合成的文本信息;对所述文本信息进行分词处理,得到至少一个语义单元;根据预设数据传输长度和所述语义单元生成待传输语音片段,所述待传输语音片段是若干个语义单元对应的合成语音,且所述若干个语义单元对应的合成语音的数据长度之和不大于所述预设数据传输长度;发送所述待传输语音片段。

[0010] 另一方面,一种应用于终端设备的合成语音的传输方法,包括:向云端服务器发送语音合成请求,所述语音合成请求由待合成的文本信息生成,以使所述云端服务器通过响应所述语音合成请求对所述文本信息进行语音合成;接收所述云端服务器返回的传输语音片段,其中,所述传输语音片段是若干个语义单元对应的合成语音,且所述若干个语义单元对应的合成语音的数据长度之和不大于预设数据传输长度;播报所述传输语音片段。

[0011] 再一方面,本公开提供一种云端服务器,所述云端服务器包括:信息接收模块,用于接收待合成的文本信息;分词处理模块,用于对所述文本信息进行分词处理,得到至少一个语义单元;判断模块,用于判断所述文本信息对应的合成语音的数据长度是否大于预设数据传输长度;若为是,则通知语音片段划分模块;所述语音片段划分模块,用于根据所述预设数据传输长度和语义单元,将所述文本信息对应的合成语音划分为至少两个待传输语音片段,所述待传输语音片段是若干个语义单元对应的合成语音;发送模块,用于发送所述待传输语音片段。

[0012] 再一方面,本公开提供一种云端服务器,所述云端服务器包括:信息接收模块,用于接收待合成的文本信息;分词处理模块,用于对所述文本信息进行分词处理,得到至少一个语义单元;语音片段生成模块,用于根据预设数据传输长度和所述语义单元生成待传输语音片段,所述待传输语音片段是若干个语义单元对应的合成语音,且所述若干个语义单元对应的合成语音的数据长度之和不大于所述预设数据传输长度;发送模块,用于发送所述待传输语音片段。

[0013] 再一方面,本公开提供一种终端设备,所述终端设备包括:发送模块,用于向云端服务器发送语音合成请求,所述语音合成请求由待合成的文本信息生成,以使所述云端服务器通过响应所述语音合成请求对所述文本信息进行语音合成;接收模块,用于接收所述云端服务器返回的传输语音片段,其中,所述传输语音片段是若干个语义单元对应的合成语音,且所述若干个语义单元对应的合成语音的数据长度之和不大于预设数据传输长度;语音播报模块,用于播报所述传输语音片段。

[0014] 与现有技术相比,本公开具有以下有益效果:

[0015] 通过对待合成的文本信息进行分词处理,得到若干个语义单元,并通过预设数据传输长度和语义单元对文本信息对应的合成语音进行划分,使得划分得到的待传输语音片段是由若干个语义单元对应的合成语音组成的,进而传输该待传输语音片段至终端设备。可以理解,由于待传输语音片段是由若干个语义单元对应的合成语音组成的,因此,无论网络环境是否异常,该待传输的语音段都将保持文本信息原有的语义结构,从而保证了经传输的合成语音的可理解性。

[0016] 应当理解的是,以上的一般描述和后文的细节描述仅是示例性和解释性的,并不能限制本公开。

附图说明

[0017] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本公开的实施例,并于说明书一起用于解释本公开的原理。

[0018] 图1是基于云端处理的语音合成过程所涉及的实施环境的示意图;

[0019] 图2是现有技术所涉及的语音合成过程的流程图;

[0020] 图2a是图2所涉及的语音合成过程中步骤330在一个实施例的流程图;

[0021] 图3是现有技术所涉及的HTS语音合成系统的示意图;

[0022] 图3a是图3所示出的HTS语音合成系统中合成声码器470的示意图;

[0023] 图4是现有技术所涉及的按照固定数据传输长度划分文本信息对应的合成语音的示意图;

[0024] 图5是根据一示例性实施例示出的一种云端服务器的框图;

[0025] 图6是根据一示例性实施例示出的一种合成语音的传输方法的流程图;

[0026] 图7是根据一示例性实施例示出的另一种合成语音的传输方法的流程图;

[0027] 图8是根据一示例性实施例示出的另一种合成语音的传输方法的流程图;

[0028] 图9是本公开所涉及的按照语义单元的发音时长划分合成语音的示意图;

[0029] 图10是图6对应实施例中步骤570在一个实施例的流程图;

[0030] 图11是图6对应实施例中步骤570在另一个实施例的流程图;

[0031] 图12是一应用场景中一种合成语音的传输方法的具体实现示意图;

[0032] 图13是根据一示例性实施例示出的另一种合成语音的传输方法的流程图;

[0033] 图14是根据一示例性实施例示出的另一种合成语音的传输方法的流程图;

[0034] 图15是图13对应实施例中步骤950在一个实施例的流程图;

[0035] 图16是根据一示例性实施例示出的另一种合成语音的传输方法的流程图;

[0036] 图17是根据一示例性实施例示出的一种合成语音的传输装置的框图;

[0037] 图18是根据一示例性实施例示出的另一种合成语音的传输装置的框图;

[0038] 图19是根据一示例性实施例示出的另一种合成语音的传输装置的框图。

[0039] 通过上述附图,已示出本公开明确的实施例,后文中将有更详细的描述,这些附图和文字描述并不是为了通过任何方式限制本公开构思的范围,而是通过参考特定实施例为本领域技术人员说明本公开的概念。

具体实施方式

[0040] 这里将详细地对示例性实施例执行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本公开相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本公开的一些方面相一致的装置和方法的例子。

[0041] 图1为基于云端处理的语音合成过程所涉及的实施环境。该实施环境包括云端服务器100和终端设备200。

[0042] 其中,云端服务器100用于对接收到的待合成的文本信息进行语音合成得到合成语音,并通过网络向终端设备200传输该合成语音。

[0043] 终端设备200用于向云端服务器100发送待合成的文本信息,并对云端服务器100返回的合成语音进行播报,以使用户掌握播报内容。该终端设备200可以是智能手机、平板电脑、掌上电脑、笔记本电脑或者设置有音频播放器的其它电子设备和嵌入式设备。

[0044] 通过云端服务器100与终端设备200之间如上所述的交互,完成文字信息转换为声音信息的语音合成过程。

[0045] 现结合图1,对现有技术所涉及的语音合成过程加以详细说明如下,如图2所示,该语音合成过程可以包括以下步骤:

[0046] 步骤310,接收由终端设备发送的待合成的文本信息。

[0047] 待合成的文本信息可以由终端设备200内部生成的,也可以是由与终端设备200相连的外部设备输入的,例如,外部设备为键盘等,本公开对待合成的文本信息的输入方式不做限定。

[0048] 在终端设备200得到待合成的文本信息之后,即可向云端服务器100发送该待合成的文本信息,以通过云端服务器100对该待合成的文本信息进行后续的语音合成。

[0049] 进一步地,终端设备200通过发送语音合成请求至云端服务器100,实现待合成的文本信息的语音合成。其中,该语音合成请求是由待合成的文本信息生成的。

[0050] 步骤330,对待合成的文本信息进行文本分析,得到文本分析结果。

[0051] 文本分析指的是模拟人对自然语言的理解过程,使云端服务器100可以在一定程度上对该待合成的文本信息进行理解,从而知道该待合成的文本信息发什么音、怎么发音以及发音的方式。此外,还能够使云端服务器100了解该待合成的文本信息中包含哪些词、短语和句子、发音时哪里需要停顿、以及停顿的时间等等。

[0052] 由此,如图2a所示,文本分析过程可以包括以下步骤:

[0053] 步骤331,对待合成的文本信息进行规范化处理。

[0054] 规范化处理是指将待合成的文本信息中不规范或者无法正常发音的字符过滤掉,例如,待合成的文本信息中出现的乱码、或者其他无法进行语音合成的语言类型等等。

[0055] 步骤333,对规范化处理的文本信息进行分词处理,得到分词文本。

[0056] 分词处理可以根据规范化处理的文本信息的上下文关系进行,还可以根据预先构建的词典模型进行。

[0057] 具体地,通过分词处理得到的分词文本包含至少一个语义单元。该语义单元指的是用户可理解的具有完整词语解释的单元,该语义单元可以由若干词、若干短语、甚至于若干句子组成。

[0058] 举例来说,规范化处理的文本信息为“云端语音合成技术,基于云端处理,将文字信息转换为声音信息。”,经分词处理后,得到的分词文本如表1所示。

[0059] 表1 分词文本

[0060]

分词文本	云端 语音 合成 技术 , 基于 云端 处理 , 将 文字 信息 转换 为 声音 信息 。
------	---

[0061] 其中,“云端”、“语音”、“合成”、“技术”等等均可视为语义单元。

[0062] 当然,在不同的应用场景中,分词文本中包含的语义单元还可以是英文串、数字串、符号串等等。

[0063] 步骤335,根据建立的韵律声学模型确定分词文本所对应的文本分析结果。

[0064] 由于分词文本包含若干语义单元,该语义单元是用户可理解的具有完整词语解释的单元,基于此,分词文本能够反映出待合成的文本信息原有的语义结构,而文本分析结果则能够在一定程度上反映出待合成的文本信息原有的韵律信息。由于语音合成时更多的是基于人特有的韵律节奏来发音的,因此,在进行语音合成之前,需要将分词文本转化为文本分析结果。

[0065] 进一步地,在确定分词文本所对应的文本分析结果之前,还需要建立语义结构所对应的韵律声学模型。

[0066] 韵律声学模型的建立过程包括:根据韵律节奏对韵律短语和重音进行预测,并通过预测结果与实际语境的相互结合来实现韵律声学参数的预测和选取,从而根据得到的韵律声学参数完成韵律声学模型的建立。

[0067] 在得到韵律声学模型之后,即可通过韵律声学模型对分词文本的韵律边界进行调整,并对调整后的分词文本进行韵律信息的标注,例如,韵律信息的标注可以包括确定调整的分词文本的发音、以及发音时的语气变换和轻重方式,从而形成分词文本对应的文本分析结果,以供后续语音合成过程中使用。

[0068] 举例来说,如上表1所示的分词文本中,“转换|为”经过韵律边界调整之后调整为“转换为”,再经过韵律信息的标注之后,其所对应的文本分析结果为“zhuan3huan4wei2”。

[0069] 步骤350,通过语音合成技术将文本分析结果合成为合成语音。

[0070] 以语音合成技术采用HTS语音合成系统为例,对文本分析结果合成为合成语音的语音合成原理加以说明如下。

[0071] 如图3所示,HTS语音合成系统400包括模型训练部分和语音合成部分。其中,模型训练部分包括训练语料库410、激励参数提取单元420、频谱参数提取单元430和HMM训练单元440。语音合成部分包括文本分析及状态转换单元450、合成参数生成器460和合成声码器470。

[0072] 模型训练部分:在进行隐马尔可夫模型(HMM模型)训练之前,一方面,需要对训练语料库410中存储的训练语料进行时间标注,以生成具有时长信息的标注序列(例如语音帧);另一方面,需要由训练语料中提取出语音合成所需要的参数,该参数包括激励参数、频谱参数和状态持续时间参数。

[0073] 进一步地,通过激励参数提取单元420对训练语料进行基频特征的提取,形成激励信息;通过频谱参数提取单元430对训练语料进行梅尔频率倒谱系数(MFCC)的提取,形成频谱参数;状态持续时间参数则是在隐马尔可夫模型训练过程中生成的。

[0074] 之后,将标注序列、激励参数和频谱参数输入至HMM训练单元440进行隐马尔可夫模型的训练,从而为每一个标注序列(例如每一个语音帧)建立对应的隐马尔可夫模型,以供后续语音合成时使用。

[0075] 语音合成部分:待合成的文本信息由文本分析及状态转换单元450进行文本分析和状态转换,即待合成的文本信息经文本分析得到文本分析结果,文本分析结果再经状态转换形成对应的隐马尔可夫模型中的状态序列。

[0076] 然后,将状态序列输入至合成参数生成器460,基于状态序列所包含的状态持续时间参数,通过参数生成算法计算出该状态序列所对应的激励参数和频谱参数。

[0077] 进一步地,如图3a所示,合成声码器470包括滤波器参数校正器471、激励信号生成器473和MLSA滤波器475。

[0078] 其中,滤波器参数校正器471用于根据状态序列对应的频谱参数校正MLSA滤波器475的系数,从而使MLSA滤波器475能够模仿人体口腔和声道特征。

[0079] 激励信号生成器473用于根据状态序列对应的激励参数来判断清、浊音从而产生不同的激励信号。若判断为浊音,则产生以激励参数周期为周期的脉冲序列作为激励信号;若判断为清音,则产生高斯白噪声序列作为激励信号。

[0080] 具体而言,在计算得到状态序列对应的激励参数和频谱参数之后,将频谱参数输入滤波器参数校正器471以对MLSA滤波器475的系数进行校正,将激励参数输入激励信号生成器473产生激励信号,进而以该激励信号作为激励源通过校正后的MLSA滤波器475,即可合成得到该状态序列所对应的语音。

[0081] 值得一提的是,文本分析结果经状态转换可能形成若干状态序列,每一状态序列均能够合成得到对应的语音,相应地,合成语音将由若干语音组成,使得合成语音具有一定的时长。

[0082] 当然,在其他应用场景中,也可以采用其余语音合成系统进行语音合成,本公开并不对此加以限制。

[0083] 待完成上述步骤,即完成基于云端处理的语音合成过程。

[0084] 由上可知,待合成的文本信息合成为合成语音需要消耗一定的时间,若云端服务器100待合成语音全部合成完毕才将合成语音返回至终端设备200,则可能导致语音合成过程耗时过长,而如果云端服务器100按照固定数据传输长度将文本信息对应的合成语音划分为待传输语音片段进行传输,虽然有效地缩短了语音合成过程的时长,但是由于网络环境的影响,可能导致待传输语音片段之间不连续,而破坏了待合成的文本信息原有的语义结构,进而导致用户无法理解终端设备所播报的内容。

[0085] 举例来说,图4是现有技术所涉及的按照固定数据传输长度划分文本信息对应的合成语音的示意图。其中,合成语音所对应的文本信息的内容为“云端语音合成技术,基于云端处理,将文字信息转换为声音信息。”。

[0086] 如图4所示,现有技术中,按照固定数据传输长度N对文本信息对应的合成语音进行待传输语音片段的划分,将得到7个待传输语音片段,该7个待传输语音片段所对应的文本信息的内容分别为:“云端语音合”、“成技术,基于”、“云端处理”、“,将文字”、“信息转换”、“为声音信”、“息。”。

[0087] 由此可知,在网络环境异常时,由于待传输语音片段之间不连续,将导致待传输语音片段所对应的文本信息的内容中断,例如,“云端语音合”、“成技术,基于”之间的停顿即不符合待合成的文本信息原有的语义结构,而导致合成语音的可理解性大大降低,降低了用户体验。

[0088] 因此,为了在网络环境异常时提高经传输的合成语音的可理解性,特提出了一种合成语音的传输方法,该种合成语音的传输方法适用于图1所示实施环境中的云端服务器100。

[0089] 图5是根据一示例性实施例示出的一种云端服务器100的框图。该硬件结构只是一个适用本公开的示例,不能认为是对本公开的使用范围的任何限制,也不能解释为本公开

需要依赖于该云端服务器100。

[0090] 该云端服务器100可因配置或者性能的不同而产生较大的差异,如图2所示,云端服务器100包括:电源110、接口130、至少一存储介质150、以及至少一中央处理器(CPU, Central Processing Units)170。

[0091] 其中,电源110用于为云端服务器100上的各硬件设备提供工作电压。

[0092] 接口130包括至少一有线或无线网络接口131、至少一串并转换接口133、至少一输入输出接口135以及至少一USB接口137等,用于与外部设备通信。

[0093] 存储介质150作为资源存储的载体,可以是随机存储介质、磁盘或者光盘等,其上所存储的资源包括操作系统151、应用程序153及数据155等,存储方式可以是短暂存储或者永久存储。其中,操作系统151用于管理与控制云端服务器100上的各硬件设备以及应用程序153,以实现中央处理器170对海量数据155的计算与处理,其可以是Windows Server™、Mac OS X™、Unix™、Linux™、FreeBSD™等。应用程序153是基于操作系统151之上完成至少一项特定工作的计算机程序,其可以包括至少一模块(图示未示出),每个模块都可以分别包含有对云端服务器100的一系列操作指令。数据155可以是存储于磁盘中的照片、图片等等。

[0094] 中央处理器170可以包括一个或多个以上的处理器,并设置为通过总线与存储介质150通信,用于运算与处理存储介质150中的海量数据155。

[0095] 如上面所描述的,适用本公开各示例性实施例的云端服务器100可以用于实现合成语音的动态长度传输,即通过中央处理器170读取存储介质150中存储的一系列操作指令的形式,按照预设数据传输长度和语义单元对文本信息对应的合成语音进行待传输语音片段的划分,并传输该待传输语音片段至终端设备200,以通过终端设备200进行语音播报,使得用户掌握播报内容。

[0096] 此外,通过硬件电路或者硬件电路结合软件指令也能同样实现本公开,因此,实现本公开并不限于任何特定硬件电路、软件以及两者的组合。

[0097] 请参阅图6,在一示例性实施例中,一种合成语音的传输方法适用于图1所示实施环境中的云端服务器100,该种合成语音的传输方法可以由云端服务器100执行,可以包括以下步骤:

[0098] 步骤510,接收待合成的文本信息。

[0099] 如前所述,待合成的文本信息可以由终端设备内部生成的,也可以是由与终端设备相连的外部设备输入的,例如,外部设备为键盘等。

[0100] 在终端设备得到待合成的文本信息之后,即可向云端服务器发送该待合成的文本信息,以通过云端服务器对该待合成的文本信息进行后续的语音合成。

[0101] 进一步地,终端设备通过发送语音合成请求至云端服务器,实现待合成的文本信息的语音合成。其中,该语音合成请求是由待合成的文本信息生成的。

[0102] 步骤530,对文本信息进行分词处理,得到至少一个语义单元。

[0103] 如前所述,通过文本信息的分词处理得到的分词文本包含至少一个语义单元,该语义单元指的是用户可理解的具有完整词语解释的单元,该语义单元可以由若干词、若干短语、甚至于若干句子组成。例如,“云端”、“语音”、“合成”、“技术”等词均属于分词文本中所包含的语义单元。

[0104] 当然,在不同的应用场景中,分词文本中包含的语义单元还可以是英文串、数字串、符号串等等。

[0105] 步骤550,判断文本信息对应的合成语音的数据长度是否大于预设数据传输长度。

[0106] 可以理解,若文本信息对应的合成语音的数据长度未超过预设数据传输长度,则表示云端服务器仅需要进行一次传输,即可将合成语音全部发送至终端设备。此时,云端服务器可以直接进行文本信息对应的合成语音的传输,而无需对文本信息对应的合成语音进行传输处理。

[0107] 基于此,云端服务器将通过判断文本信息对应的合成语音的数据长度是否大于预设数据传输长度,来判断是否对文本信息对应的合成语音进行传输处理。

[0108] 在判断到文本信息对应的合成语音的数据长度大于预设数据传输长度时,则进入步骤570,以对文本信息对应的合成语音进行传输处理。

[0109] 反之,在判断到文本信息对应的合成语音的数据长度不大于预设数据传输长度时,则进入步骤590,直接发送文本信息对应的合成语音,即文本信息对应的合成语音即为待传输语音片段。

[0110] 步骤570,根据预设数据传输长度和语义单元,将文本信息对应的合成语音划分为至少两个待传输语音片段。

[0111] 本实施例中,对文本信息对应的合成语音进行的传输处理是通过对文本信息对应的合成语音进行待传输语音片段的划分完成。

[0112] 所述划分可以根据语义单元的数量进行,也可以根据语义单元对应的合成语音的数据长度进行。

[0113] 由于每个语义单元对应的合成语音的数据长度各不相同,两个语义单元和三个语义单元各自对应的合成语音的数据长度可能非常接近。若仅根据语义单元的数量进行文本信息对应的合成语音的划分,则可能导致划分得到的待传输语音片段的数据长度差异太大,使得终端设备进行语音播报时长时短而导致用户体验差。

[0114] 因此,较优地,为了保证划分所得到的待传输语音片段的数据长度大致相同,云端服务器将结合预设数据传输长度和语义单元进行文本信息对应的合成语音的划分,即在待传输语音片段的数据长度不超过预设数据传输长度的前提下,使待传输语音片段由若干个语义单元对应的合成语音组成。例如,待传输语音片段既可能由两个语义单元对应的合成语音组成,也可能由三个语义单元对应的合成语音组成,甚至由更多个语义单元对应的合成语音组成,以使终端设备进行语音播报的时长大致相同,进而提高用户体验

[0115] 需要说明的是,本实施例中,云端服务器是在文本信息已经合成为对应的合成语音之后,才开始合成语音的传输,以满足对语音合成的质量要求较高的应用场景。

[0116] 可以理解,云端服务器将首先存储文本信息对应的合成语音,待完成文信息对应的合成语音的划分之后,才开始传输划分得到的待传输语音片段。

[0117] 步骤590,发送待传输语音片段。

[0118] 终端设备在接收到待传输语音片段,即根据该待传输语音片段进行语音播报。

[0119] 由于该待传输语音片段是由若干个语义单元对应的合成语音组成的,因此,每一次播报内容都是用户所能理解的。例如,待传输语音片段所对应的文本信息的内容是“云端语音”。

[0120] 通过如上所述的过程,实现了合成语音的动态长度传输,即待传输语音片段的数据长度并不是固定长度,而是由组成其的若干个语义单元对应的合成语音的数据长度决定的,由于语义单元遵循待合成的文本信息原有的语义结构,从而保证即使网络环境异常导致若干待传输语音片段之间不连续,也不会破坏待合成的文本信息原有的语义结构,以此有效地提高了经传输的合成语音的可理解性,提高了用户体验。

[0121] 请参阅图7,在一示例性实施例中,步骤550之前,如上所述的方法还可以包括以下步骤:

[0122] 步骤610,监测网络状态。

[0123] 步骤630,根据监测到的网络状态调整预设数据传输长度。

[0124] 预设数据传输长度即为前述PCM数据传输方法进行合成语音传输时所设置的固定数据传输长度。

[0125] 如前所述,该预设数据传输长度在网络环境正常时,不会影响合成语音的传输,即终端设备能够及时接收到由合成语音划分为的固定长度的若干待传输语音片段,并播报这些待传输语音片段。若网络环境异常,则可能导致终端设备接收到的固定长度的若干待传输语音片段不连续,即存在随机停顿,而可能破坏了待合成的文本信息原有的语义结构,进而导致用户无法正常理解终端设备所播报的内容。

[0126] 为此,本实施例中,将进一步地结合当前的网络环境对该预设数据传输长度进行调整,以此保证终端设备进行语音播报的流畅性。

[0127] 较优地,当前的网络环境通过监测网络状态实现。该监测可以是对终端设备的当前网速进行监测,也可以是对终端设备的当前连接状态进行监测,进而根据监测结果调整预设数据传输长度

[0128] 举例来说,通过网络状态监测得到终端设备的当前网速为S,而合成语音所需要的网速设置为M,则合成语音的预设数据传输长度可以按照下述公式进行调整:

$$[0129] \quad N' = \frac{S}{M} \times N,$$

[0130] 其中, N' 为调整后的预设数据传输长度, N 为预设数据传输长度。

[0131] 应当理解,当S小于M时, N' 小于N,表示调整后的预设数据传输长度 N' 小于预设数据传输长度N,以此适应网速较差的网络环境,即在网速较差时降低单位时间内合成语音的传输数据量。同理,在网速较好时则提高单位时间内合成语音的传输数据量,以此保证终端设备进行语音播报的流畅性。

[0132] 进一步地,为预设数据传输长度N设置一最小值 N_{\min} 。当 $N' < N_{\min}$ 时,令 $N' = N_{\min}$ 。也就是说,若调整后的预设数据传输长度 N' 比最小的预设数据传输长度 N_{\min} 还小,则以最小的预设数据传输长度 N_{\min} 作为预设数据传输长度N,以此避免云端服务器与终端设备之间的交互过于频繁,从而有效地提高云端服务器的处理效率。

[0133] 进一步地,在根据网络环境对预设数据传输长度进行调整之后,步骤550中的判断即是基于调整后的预设数据传输长度进行的,以此动态地适应网络环境,从而有利于后续的合成语音传输。

[0134] 通过如上所述的过程,结合当前的网络环境,实现了对合成语音的预设数据传输长度的动态调整,使得合成语音在网络异常时能够以较小的传输长度进行传输,进而有利

于保证待传输语音片段之间传输的连续性,以此保证终端设备能够不间断地播报接收到的待传输语音片段,从而有利于提高经传输的合成语音的可理解性。

[0135] 请参阅图8,在一示例性实施例中,步骤550之前,如上所述的方法可以包括以下步骤:

[0136] 步骤710,根据汉语发音时长计算文本信息包括的各语义单元的发音时长。

[0137] 如前所述,语义单元可以包括若干词、若干短语、甚至是若干句子,而无论上述何种形式的语义单元均是由语法结构中的基本单位——词构成的。

[0138] 相应地,词的发音时长与汉语发音时长相关,即与汉语的声母、韵母的发音时长相关。可以理解,各词之间具有不同的发音时长,如图9所示,双音节语素构成的词“云端”、“语音”、“合成”、“技术”对应的双音节分别为“yunduan”、“yuyin”、“hecheng”、“jishu”,其相应的发音时长分别为 l_0 、 l_1 、 l_2 、 l_3 。因此,通过汉语发音时长即可计算得到各语义单元的发音时长。

[0139] 步骤730,根据文本信息包括的各语义单元的发音时长之和,得到文本信息的发音时长。

[0140] 由于文本信息包括若干个语义单元,因此,在计算得到文本信息包括的各语义单元的发音时长之后,即可进一步地计算得到文本信息包括的所有语义单元的发音时长之和,亦即文本信息的发音时长。

[0141] 如图9所示,文本信息的发音时长 $l = l_0 + l_1 + l_2 + l_3 + \dots + l_{i-2} + l_{i-1}$, $i = 16$ 。

[0142] 步骤750,根据文本信息的发音时长,确定文本信息对应的合成语音的数据长度。

[0143] 由于合成语音进行传输时,是以数据包的形式进行传输的,因此,在得到文本信息的发音时长之后,需要对其进行数据量转化,即将文本信息的发音时长转化为其所对应的合成语音的数据长度,对于上述转化过程,属于现有技术的范畴,本发明实施例不做限定。

[0144] 应当理解,若文本信息的发音时长较长,则其对应的合成语音的数据长度较长,反之,若文本信息的发音时长较短,则其对应的合成语音的数据长度也较短。

[0145] 在确定文本信息对应的合成语音的数据长度之后,云端服务器即可根据该文本信息对应的合成语音的数据长度,判断后续是否需要对该文本信息对应的合成语音进行待传输语音片段的划分。

[0146] 如前所述,为了避免终端设备接收到的待传输语音片段的数据长度差异太大,使得语音播报时长时短而导致用户的体验差,云端服务器将结合预设数据传输长度和语义单元进行文本信息对应的合成语音的划分,即在待传输语音片段的数据长度不超过预设数据传输长度的前提下,使待传输语音片段由若干个语义单元对应的合成语音组成。

[0147] 进一步地,对文本信息对应的合成语音的划分可以有两种方案:第一种,通过对语义单元对应的合成语音进行组合,使其形成数据长度不超过预设数据传输长度的待传输语音片段;第二种,由文本信息对应的合成语音中剔除若干语义单元对应的合成语音,以使剩下的语义单元所对应的合成语音组成数据长度不超过预设数据传输长度的待传输语音片段。

[0148] 请参阅图10,在一示例性实施例中,对文本信息对应的合成语音的划分采取上述第一种方案,相应地,步骤570可以包括以下步骤:

[0149] 步骤571,判断文本信息中第一个语义单元对应的合成语音的数据长度是否大于

预设数据传输长度。

[0150] 若第一个语义单元对应的合成语音的数据长度不大于预设数据传输长度,则进入步骤572,将该第一个语义单元和第二个语义单元分别对应的合成语音的数据长度累加,得到第一数据长度累加和。

[0151] 在得到第一数据长度累加和之后,进入步骤573,进一步判断该第一数据长度累加和是否大于预设数据传输长度。

[0152] 若判断到该第一数据长度累加和大于预设数据传输长度,基于待传输语音片段的数据长度不超过预设数据传输长度的原则,则进入步骤574,以该第一个语义单元对应的合成语音作为待传输语音片段。

[0153] 反之,若判断到该第一数据长度累加和不超过预设数据传输长度,则进入步骤575,继续对文本信息中的其余语义单元对应的合成语音的数据长度进行累加判断,直至所有语义单元对应的合成语音的数据长度均完成累加判断。

[0154] 举例来说,将第一个语义单元、第二个语义单元和第三个语义单元分别对应的合成语音的数据长度累加,得到第二数据长度累加和。

[0155] 在得到第二数据长度累加和之后,进一步判断该第二数据长度累加和是否大于预设数据传输长度。

[0156] 若判断到该第二数据长度累加和大于预设数据传输长度,基于待传输语音片段的数据长度不超过预设数据传输长度的原则,则以第一个语义单元和第二个语义单元所对应的合成语音作为待传输语音片段。

[0157] 以此类推,直至所有语义单元所对应的合成语音均作为待传输语音片段的一部分,完成合成语音的传输。

[0158] 具体而言,如图9所示,如前所述,各语义单元的发音时长为 l_i , ($i=0\sim 16$),文本信息的发音时长为 $l = l_0 + l_1 + l_2 + l_3 + \dots + l_{i-2} + l_{i-1}$, $i=16$ 。

[0159] 相应地,令各语义单元对应的合成语音的数据长度为 L_i , ($i=0\sim 16$),文本信息对应的合成语音的数据长度为 $L = L_0 + L_1 + L_2 + L_3 + \dots + L_{i-2} + L_{i-1}$, $i=16$,预设数据传输长度为 N' 。

[0160] 当 $L > N'$ 时,云端服务器将对文本信息对应的合成语音进行待传输语音片段的划分,以通过多次传输将文本信息对应的合成语音传输至终端设备。

[0161] 第一次划分时,若 $L_0 + L_1 + L_2 > N'$ 且 $L_0 + L_1 < N'$,即文本信息中第一个、第二个语义单元分别对应的合成语音的数据长度的数据长度累加和未超过预设数据传输长度,而前三个语义单元分别对应的合成语音的数据长度的数据长度累加和超过预设数据传输长度,则根据比较结果得到第一个待传输语音片段的数据长度为: $N'_0 = L_0 + L_1$,即以第一个、第二个语义单元所对应的合成语音作为待传输语音片段。

[0162] 第二次划分时,若 $L_2 + L_3 + L_4 + L_5 > N'$ 且 $L_2 + L_3 + L_4 < N'$,即文本信息中第三个、第四个、第五个语义单元分别对应的合成语音的数据长度的数据长度累加和未超过预设数据传输长度,而第三个、第四个、第五个、第六个语义单元分别对应的合成语音的数据长度的数据长度累加和超过预设数据传输长度,则根据比较结果得到第二个待传输语音片段的数据长度为: $N'_1 = L_2 + L_3 + L_4$,即以第三个、第四个、第五个语义单元所对应的合成语音作为待传输语音片段。

[0163] 以此类推,直至所有语义单元所对应的合成语音均作为待传输语音片段的一部分,完成合成语音的传输。

[0164] 请参阅图11,在另一示例性实施例中,对文本信息对应的合成语音的划分采取上述第二种方案,相应地,步骤570可以包括以下步骤:

[0165] 步骤576,将文本信息对应的合成语音的数据长度减去倒数第一个语义单元对应的合成语音的数据长度,得到第一数据长度差值。

[0166] 在得到第一数据长度差值之后,进入步骤577,判断该第一数据长度差值是否大于预设数据传输长度。

[0167] 若判断到该第一数据长度差值不大于预设数据传输长度,则进入步骤578,将倒数第一个语义单元之前的所有语义单元对应的合成语音作为待传输语音片段。

[0168] 反之,若判断到该第一数据长度差值大于预设数据传输长度,则进入步骤579,基于该第一数据长度继续对文本信息中其余语音单元对应的合成语音的数据长度进行相减判断,直至所有语义单元对应的合成语音的数据长度均完成相减判断。

[0169] 举例来说,将该第一数据长度差值减去倒数第二个语义单元对应的合成语音的数据长度,得到第二数据长度差值。

[0170] 在得到第二数据长度差值之后,进一步判断该第二数据长度差值是否大于预设数据传输长度。

[0171] 若判断到该第二数据长度差值不大于预设数据传输长度,则将倒数第二个语义单元之前的所有语义单元对应的合成语音作为待传输语音片段。

[0172] 以此类推,直至所有语义单元所对应的合成语音作为待传输语音片段的一部分,完成合成语音的传输。

[0173] 具体而言,如图9所示,如前所述,各语义单元的发音时长为 l_i , ($i=0\sim 16$),文本信息的发音时长为 $l = l_0 + l_1 + l_2 + l_3 + \dots + l_{i-2} + l_{i-1}$, $i=16$ 。

[0174] 相应地,令各语义单元对应的合成语音的数据长度为 L_i , ($i=0\sim 16$),文本信息对应的合成语音的数据长度为 $L = L_0 + L_1 + L_2 + L_3 + \dots + L_{i-2} + L_{i-1}$, $i=16$,预设数据传输长度为 N' 。

[0175] 当 $L > N'$ 时,云端服务器将对文本信息对应的合成语音进行待传输语音片段的划分,以通过多次传输将文本信息对应的合成语音传输至终端设备。

[0176] 第一次划分时,若 $L - L_{15} - L_{14} - L_{13} > N'$ 且 $L - L_{15} - L_{14} - L_{13} - L_{12} < N'$,即文本信息对应的合成语音的数据长度减去倒数第一个、第二个、第三个、第四个语义单元对应的合成语音的数据长度的数据长度差值未超过预设数据传输长度,而减去倒数第一个、第二个、第三个语义单元对应的合成语音的数据长度的数据长度差值超过预设数据传输长度,则根据比较结果得到第一个待传输语音片段的数据长度为: $N'_0 = L - L_{15} - L_{14} - L_{13} - L_{12}$,即以倒数第四个语义单元之前的所有语义单元所对应的合成语音作为第一个待传输语音片段。

[0177] 第二次划分时,由于第一待传输语音片段已划分完毕,则文本信息对应的合成语音的数据长度更新为 $L' = L_{12} + L_{13} + L_{14} + L_{15}$,则基于该文本信息对应的合成语音的数据长度 L' 继续进行划分,若 $L' - L_{15} > N'$ 且 $L' - L_{15} - L_{14} < N'$,即文本信息对应的合成语音的数据长度减去倒数第一个、第二个语义单元对应的合成语音的数据长度的数据长度差值未超过预设数据传输长度,而减去倒数第一个语义单元对应的合成语音的数据长度的数据长度差值超过预

设数据传输长度,则根据比较结果得到第二个待传输语音片段的数据长度为: $N_1' = L' - L_{15} - L_{14}$,即以倒数第二个语义单元之前的所有语义单元所对应的合成语音作为第二个待传输语音片段。

[0178] 以此类推,直至所有语义单元所对应的合成语音均作为待传输语音片段的一部分,完成合成语音的传输。

[0179] 通过如上所述的过程,实现了合成语音的动态长度传输,即每一次待传输语音片段的数据长度都不相同,由其所包含的语义单元对应的合成语音的数据长度决定,并且传输过程中始终保持了语义单元的完整性,并不会破坏文本信息原有的语义结构,从而提高了经传输的合成语音的可理解性。

[0180] 图12是一应用场景中上述合成语音的传输方法的具体实现示意图,现结合图1所示的实施环境和图12所示的具体应用场景对本公开上述实施例中语音合成过程加以说明如下。

[0181] 终端设备200通过执行步骤801,将待合成的文本信息通过语音合成请求发送至云端服务器100。

[0182] 云端服务器100通过执行步骤802和步骤803,将接收到的待合成的文本信息合成为合成语音,并通过执行步骤804对合成语音进行存储,以利于后续合成语音的动态长度传输。

[0183] 云端服务器100通过执行步骤805,按照网络状态对合成语音的预设数据传输长度进行调整,以基于该调整后的预设数据传输长度对文本信息对应的合成语音进行待传输语音片段的划分。

[0184] 进一步地,云端服务器100通过执行步骤806进行待传输语音片段的划分,即根据调整后的预设数据传输长度以及文本信息中包含的若干语义单元,对文本信息对应的合成语音进行划分。

[0185] 在划分得到待传输语音片段之后,云端服务器100即通过执行步骤807,将待传输语音片段传输至终端设备200。

[0186] 更进一步地,若文本信息对应的合成语音未全部划分完毕,则云端服务器100将通过执行步骤808,返回步骤806,继续进行划分,直至文本信息包含的所有语义单元均作为待传输语音片段的一部分,并传输至终端设备200。

[0187] 终端设备200通过执行步骤809,利用内部设置的音频播放器对接收到的传输语音片段进行播报,以使用户根据播报内容了解待合成的文本信息的内容。

[0188] 待执行完上述步骤,即完成语音合成过程。

[0189] 在本公开实施例中,实现了合成语音的双动态长度传输,即根据网络状态和文本信息中包含的语义单元进行合成语音的动态长度传输,保证了即使网络环境异常,也不会破坏文本信息原有的语义结构,既保证了终端设备进行语音播报的流畅性,还提高了经传输的合成语音的可理解性。

[0190] 请参阅图13,在一示例性实施例中,一种合成语音的传输方法适用于图1所示实施环境中的云端服务器100,该种合成语音的传输方法可以由云端服务器100执行,可以包括以下步骤:

[0191] 步骤910,接收待合成的文本信息。

[0192] 步骤930,对文本信息进行分词处理,得到至少一个语义单元。

[0193] 步骤950,根据预设数据传输长度和语义单元生成待传输语音片段,待传输语音片段是若干个语义单元对应的合成语音,且若干个语义单元对应的合成语音的数据长度之和不大于预设数据传输长度。

[0194] 步骤970,发送待传输语音片段。

[0195] 请参阅图14,在一示例性实施例中,步骤930之前,如上所述的方法还可以包括以下步骤:

[0196] 步骤1010,根据汉语发音时长计算文本信息中第一个语义单元的发音时长。

[0197] 步骤1030,根据第一个语义单元的发音时长确定第一个语义单元对应的合成语音的数据长度。

[0198] 请参阅图15,在一示例性实施例中,步骤950可以包括以下步骤:

[0199] 步骤951,判断文本信息中第一个语义单元对应的合成语音的数据长度是否大于预设数据传输长度。

[0200] 若为否,则进入步骤953。

[0201] 步骤953,将第一个语义单元和第二个语义单元分别对应的合成语音的数据长度累加,得到第一数据长度累加和。

[0202] 步骤955,判断第一数据长度累加和是否大于预设数据传输长度。若为是,则进入步骤957。

[0203] 步骤957,以第一个语义单元对应的合成语音作为待传输语音片段。

[0204] 通过如上所述的过程,实现了合成语音的动态长度传输,即待传输语音片段的数据长度并不是固定长度,而是由组成其的若干个语义单元对应的合成语音的数据长度决定的,由于语义单元遵循待合成的文本信息原有的语义结构,从而保证即使网络环境异常导致若干待传输语音片段之间不连续,也不会破坏待合成的文本信息原有的语义结构,以此有效地提高了经传输的合成语音的可理解性,提高了用户体验。

[0205] 此外,在上述各实施例中,云端服务器是一边进行语音合成,一边对已合成好的部分合成语音进行传输的,以此有效地缩短了语音合成过程所消耗的时间,能够很好地满足对于语音合成的时间要求比较高的应用场景。

[0206] 请参阅图16,在一示例性实施例中,一种合成语音的传输方法适用于图1所示实施环境中的终端设备200,该种合成语音的传输方法可以由终端设备200执行,可以包括以下步骤:

[0207] 步骤1110,向云端服务器发送语音合成请求,语音合成请求由待合成的文本信息生成,以使云端服务器通过响应语音合成请求对文本信息进行语音合成。

[0208] 步骤1130,接收云端服务器返回的传输语音片段,其中,传输语音片段是若干个语义单元对应的合成语音,且若干个语义单元对应的合成语音的数据长度之和不大于预设数据传输长度。

[0209] 步骤1150,播报传输语音片段。

[0210] 通过如上所述的过程,有效地提高了终端设备所播报内容的可理解性,从而提高了用户体验。

[0211] 下述为本公开装置实施例,可以用于执行本公开所涉及的合成语音的传输方法。

对于本公开装置实施例中未披露的细节,请参照本公开所涉及的合成语音的传输方法实施例。

[0212] 请参阅图17,在一示例性实施例中,一种云端服务器包括但不限于:信息接收模块1210、分词处理模块1230、判断模块1250、语音片段划分模块1270和发送模块1290。

[0213] 其中,信息接收模块1210用于接收待合成的文本信息。

[0214] 分词处理模块1230用于对文本信息进行分词处理,得到至少一个语义单元。

[0215] 判断模块1250用于判断文本信息对应的合成语音的数据长度是否大于预设数据传输长度。若为是,则通知语音片段划分模块1270。

[0216] 语音片段划分模块1270用于根据预设数据传输长度和语义单元,将文本信息对应的合成语音划分为至少两个待传输语音片段,待传输语音片段是若干个语义单元对应的合成语音。

[0217] 发送模块1290用于发送待传输语音片段。

[0218] 请参阅图18,在一示例性实施例中,一种云端服务器包括但不限于:信息接收模块1310、分词处理模块1330、语音片段生成模块1350和发送模块1370。

[0219] 其中,信息接收模块1310用于接收待合成的文本信息。

[0220] 分词处理模块1330用于对文本信息进行分词处理,得到至少一个语义单元。

[0221] 语音片段生成模块1350用于根据预设数据传输长度和语义单元生成待传输语音片段,待传输语音片段是若干个语义单元对应的合成语音,且若干个语义单元对应的合成语音的数据长度之和不大于预设数据传输长度。

[0222] 发送模块1370用于发送待传输语音片段。

[0223] 请参阅图19,在一示例性实施例中,一种终端设备包括但不限于:发送模块1410、接收模块1430和语音播报模块1450。

[0224] 其中,发送模块1410用于向云端服务器发送语音合成请求,语音合成请求由待合成的文本信息生成,以使云端服务器通过响应语音合成请求对文本信息进行语音合成。

[0225] 接收模块1430用于接收云端服务器返回的传输语音片段,其中,传输语音片段是若干个语义单元对应的合成语音,且若干个语义单元对应的合成语音的数据长度之和不大于预设数据传输长度。

[0226] 语音播报模块1450用于播报传输语音片段。

[0227] 需要说明的是,上述实施例所提供的合成语音的传输装置(云端服务器、终端设备)在进行合成语音的传输时,仅以上述各功能模块的划分进行举例说明,实际应用中,可以根据需要而将上述功能分配由不同的功能模块完成,即合成语音的传输装置的内部结构将划分为不同的功能模块,以完成以上描述的全部或者部分功能。

[0228] 另外,上述实施例所提供的合成语音的传输装置与合成语音的传输方法的实施例属于同一构思,其中各个模块执行操作的具体方式已经在方法实施例中进行了详细描述,此处不再赘述。

[0229] 上述内容,仅为本公开的较佳示例性实施例,并非用于限制本公开的实施方案,本领域普通技术人员根据本公开的主要构思和精神,可以十分方便地进行相应的变通或修改,故本公开的保护范围应以权利要求书所要求的保护范围为准。

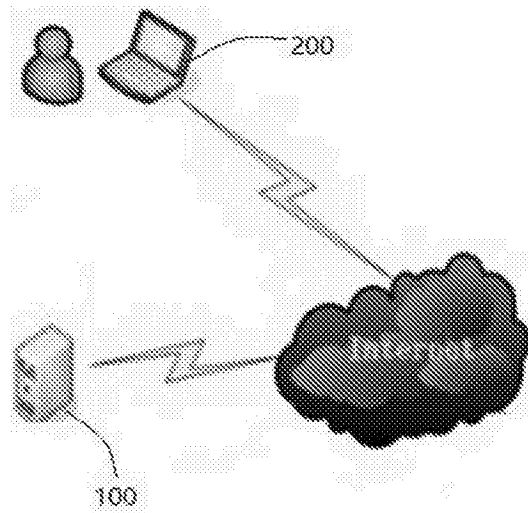


图1

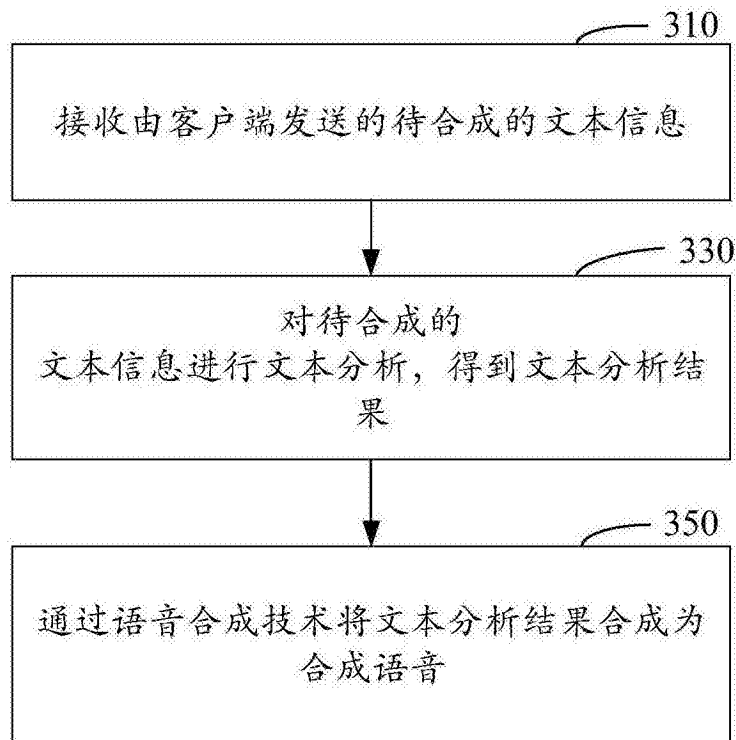


图2

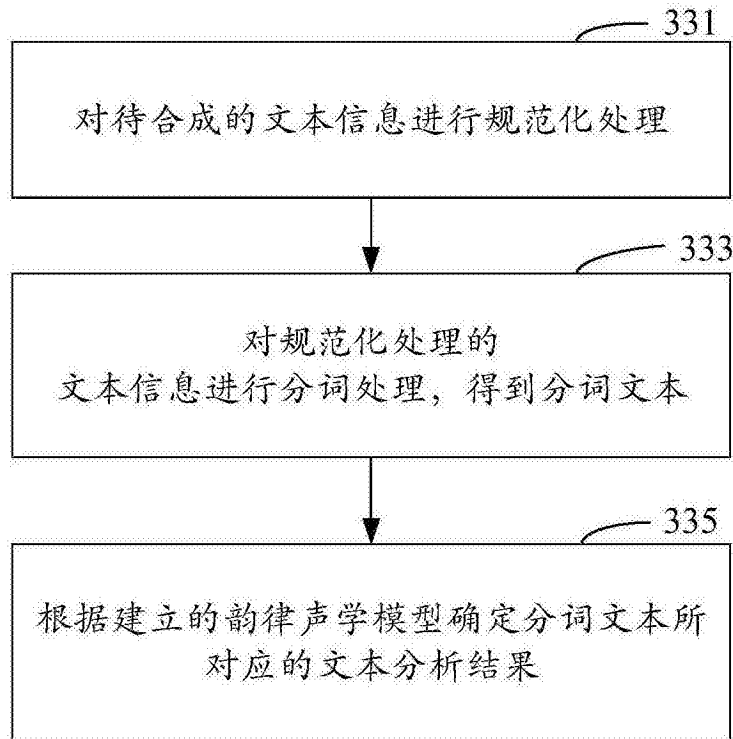


图2a

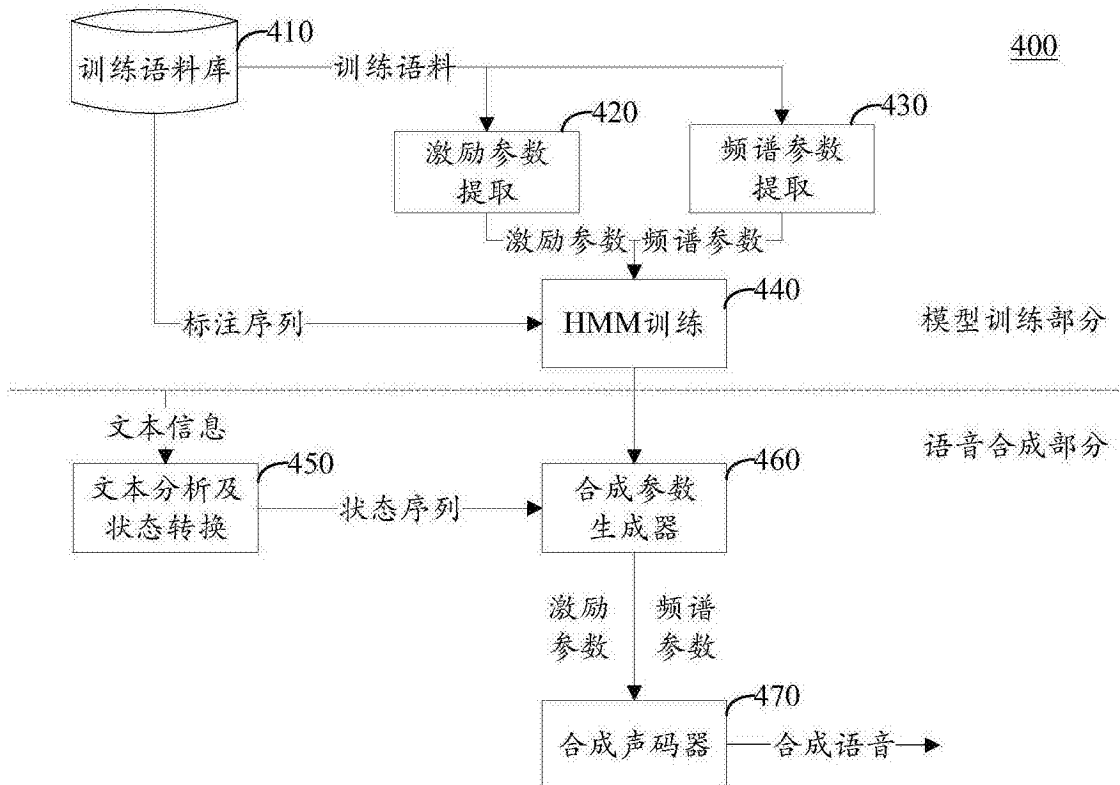


图3

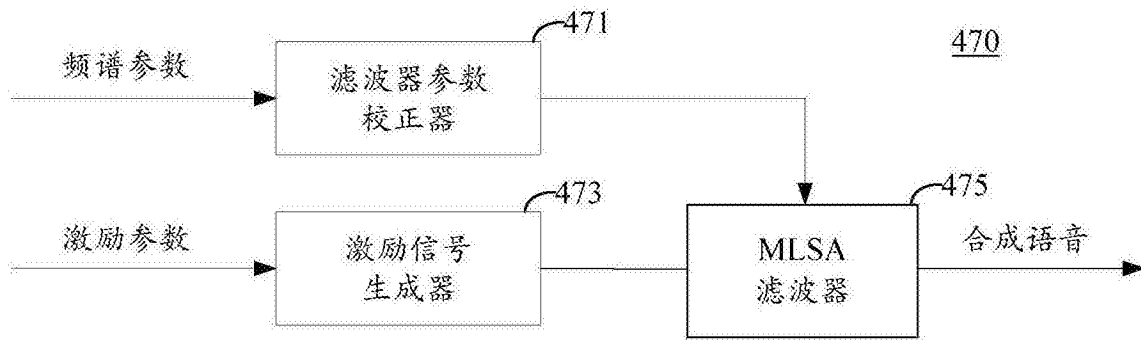


图3a

N		N			N			
yun2duan1	yu3yin1	he2 cheng2	ji4shu4	,	ji1yu2	yun2duan1	chu4li3	,
云端	语音	合成	技术	,	基于	云端	处理	,
N		N			N		N	
jiang1	wen2zi4	xin4xi1	zhuan3huan4wei2		sheng1yin1	xin4 xi1	。	
将	文字	信息	转换为		声音	信息	。	

图4

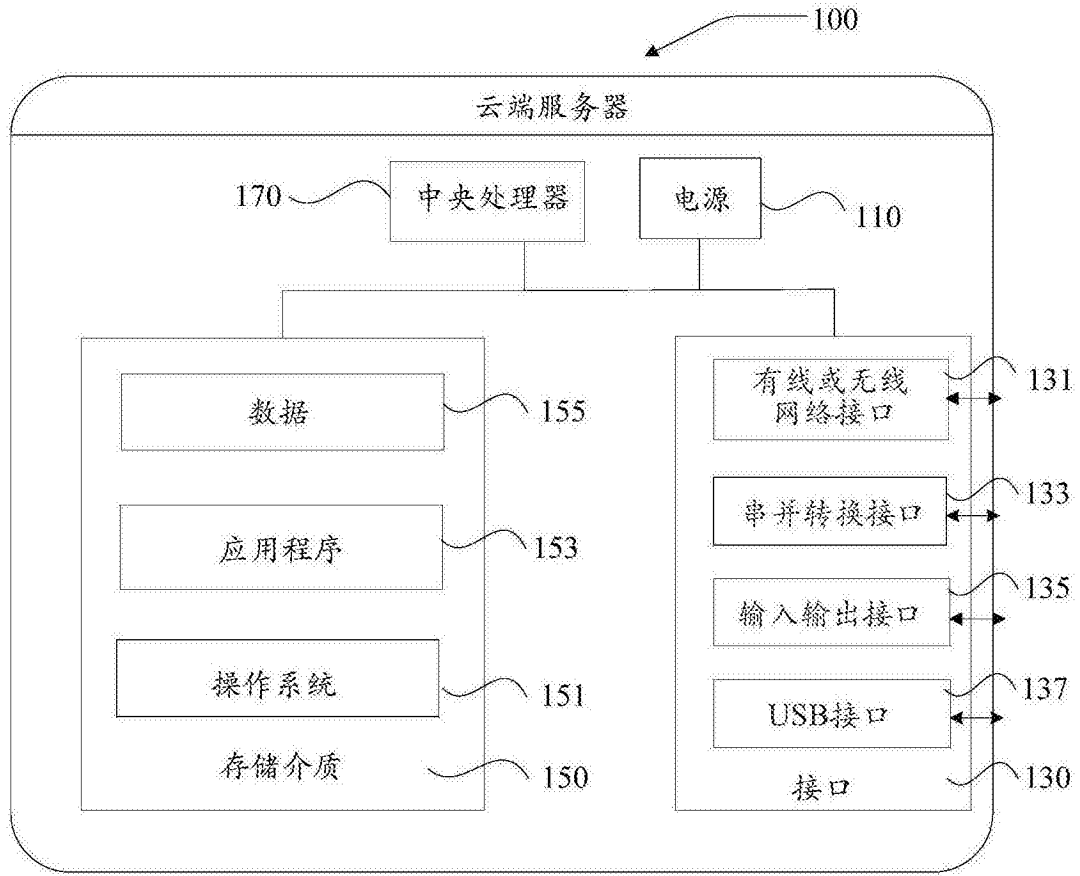


图5

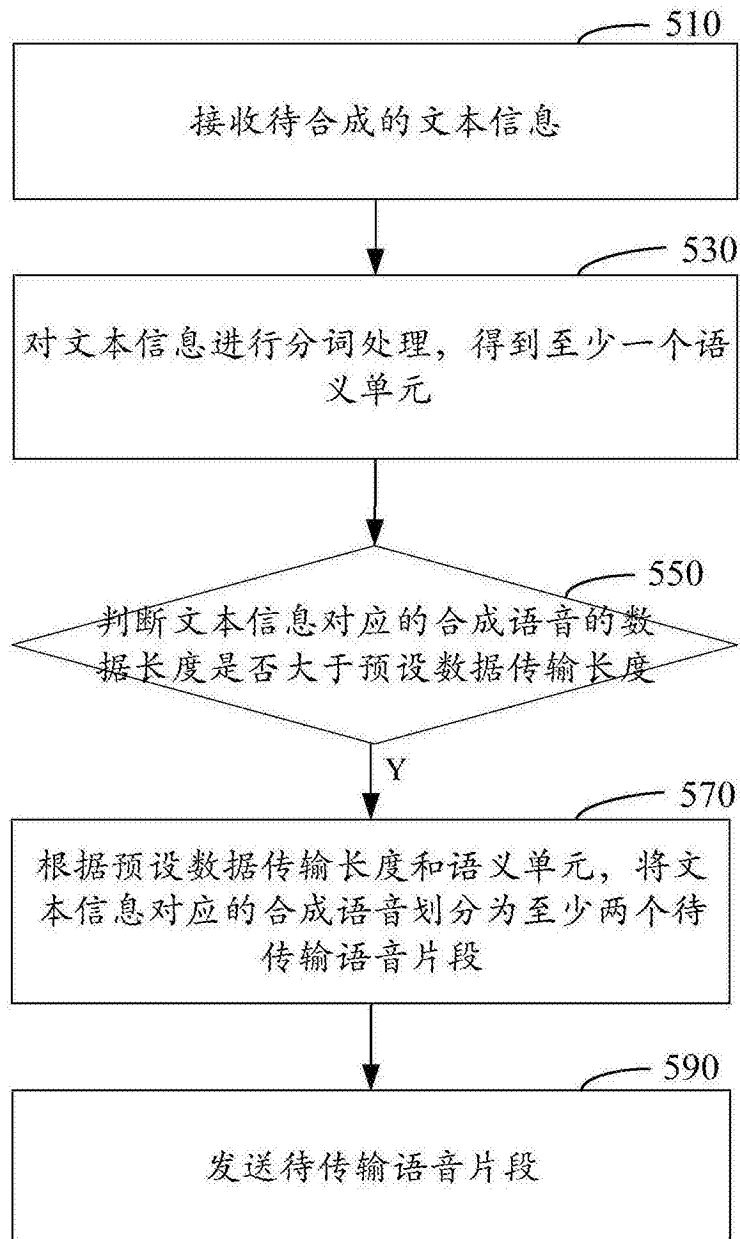


图6

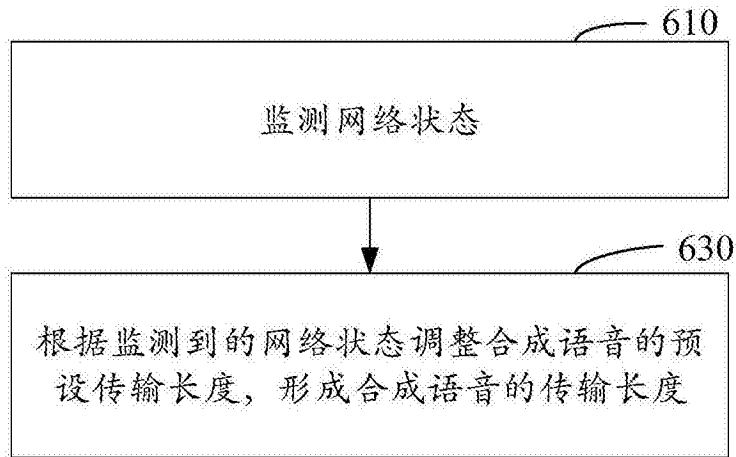


图7

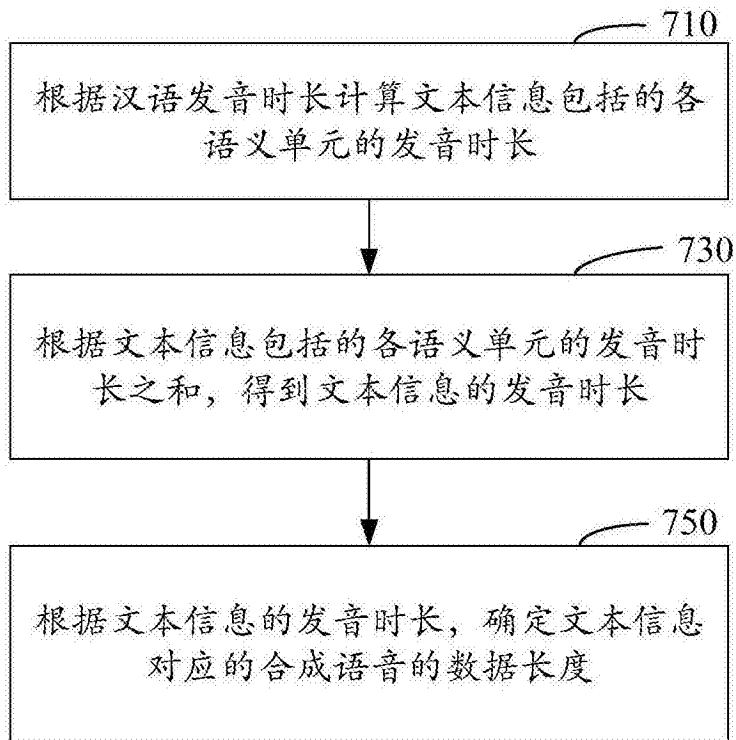


图8

l_0	l_1	l_2	l_3			
yun2duan1	yu3yin1	he2cheng2	ji4shu4	,	ji1yu2	yun2duan1	chu4li3
云端	语音	合成	技术	,	基于	云端	处理
					l_{i-2}	l_{i-1}
jiang1	wen2zi4	xin4xi1	zhuan3huan4wei2		sheng1yin1	xin4xi1	。
将	文字	信息	转换为		声音	信息	。

图9

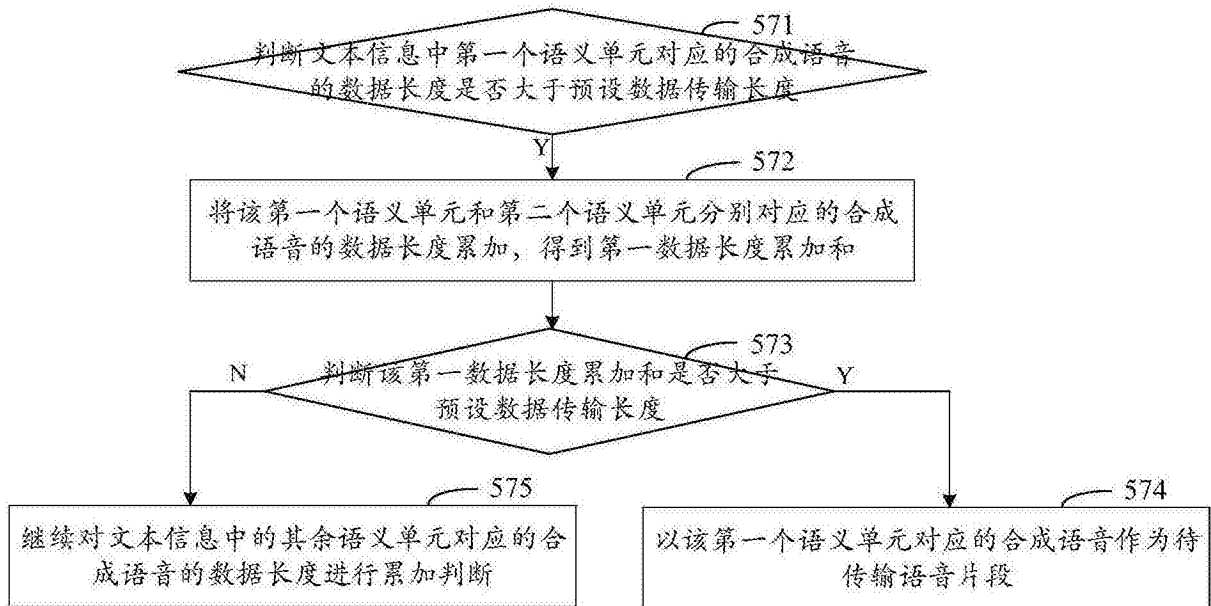


图10

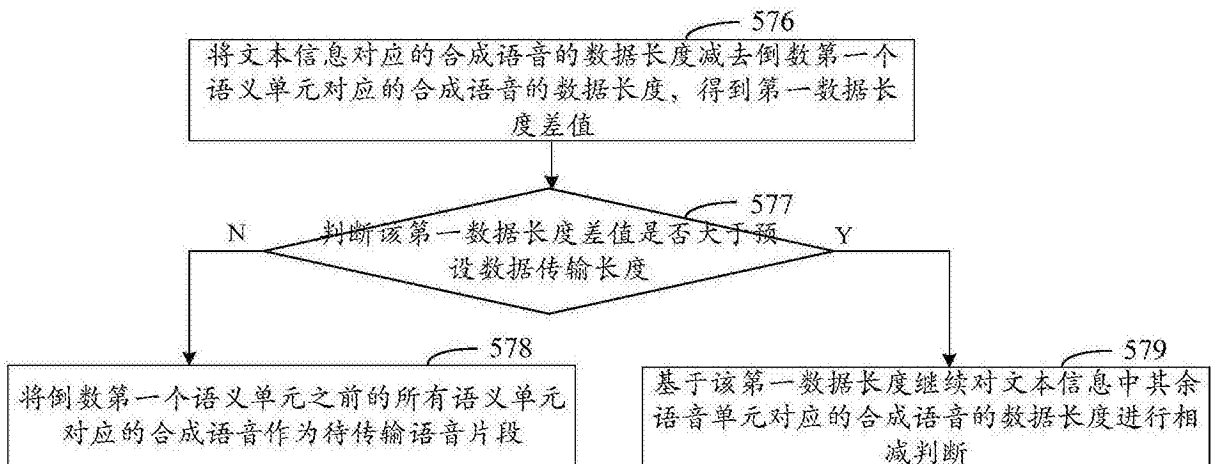


图11

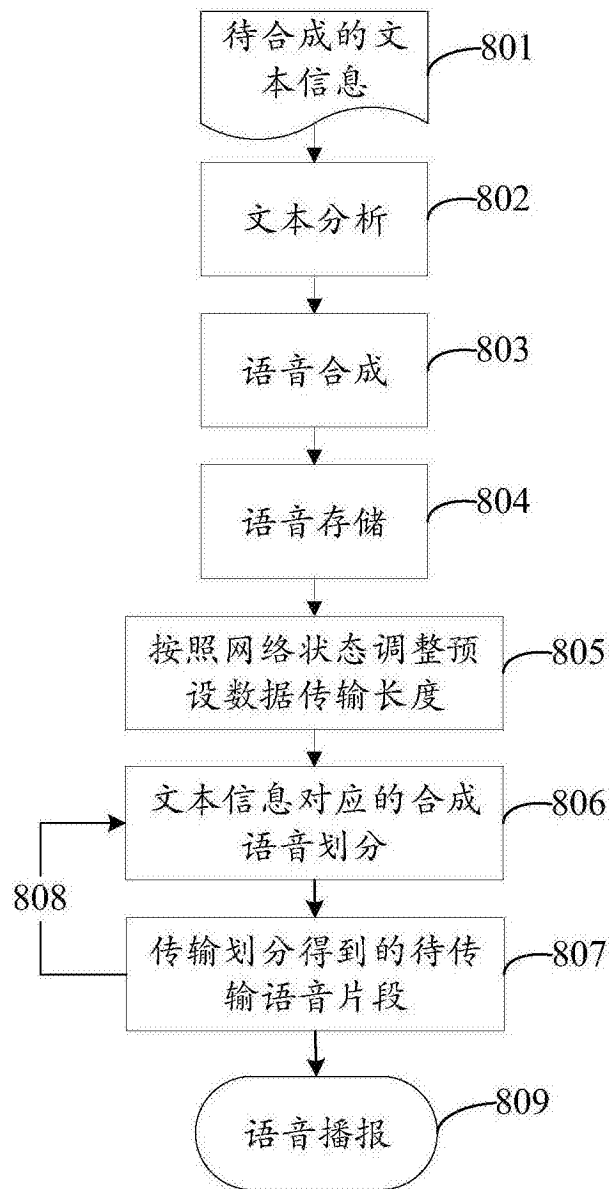


图12

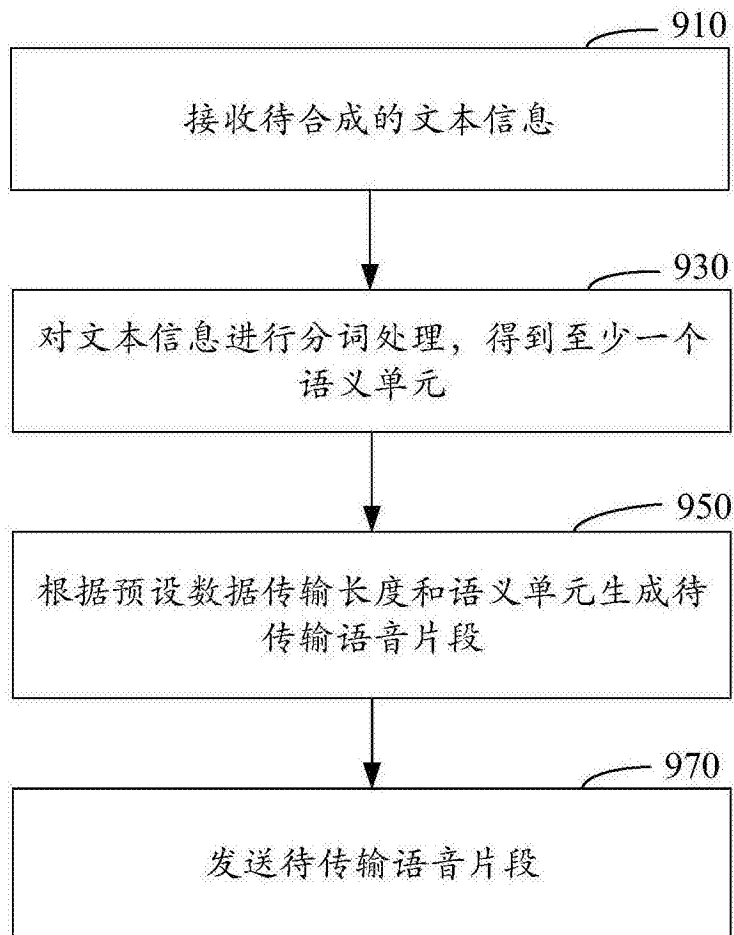


图13

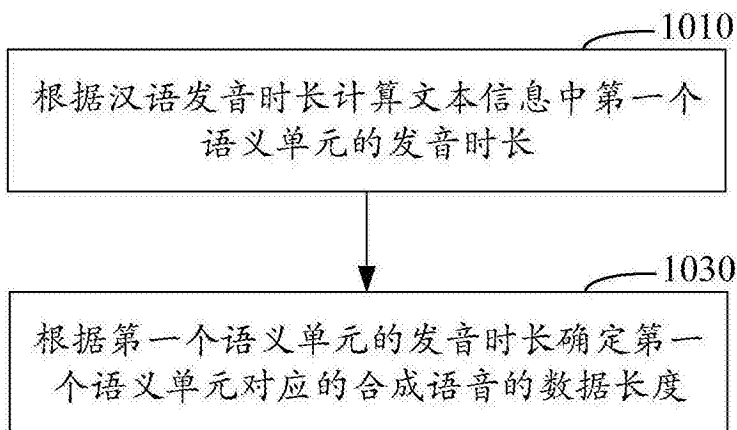


图14

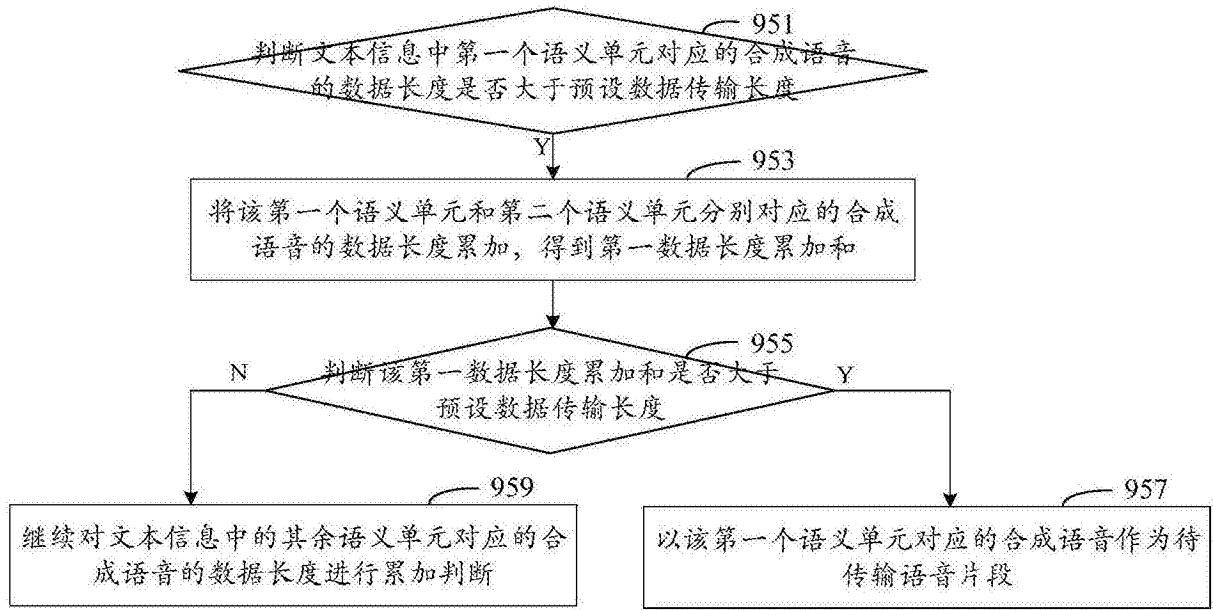


图15

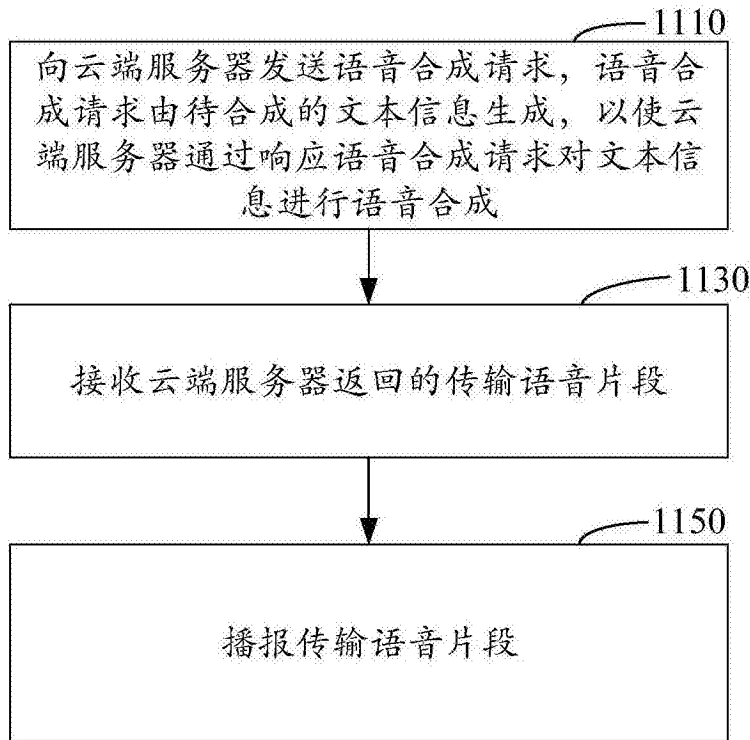


图16

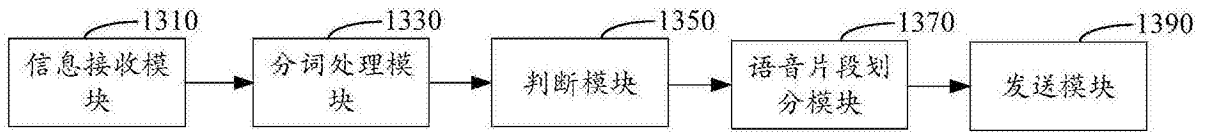


图17

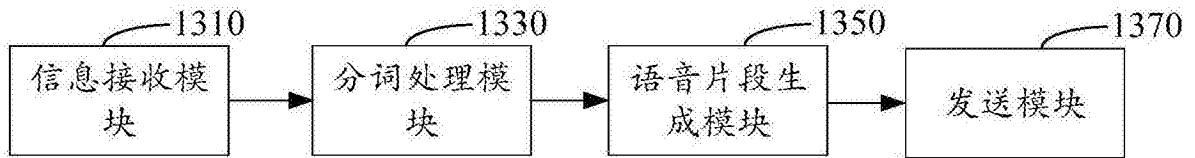


图18

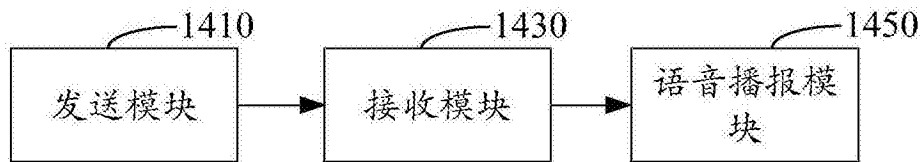


图19