

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2020 年 7 月 23 日 (23.07.2020)



(10) 国际公布号
WO 2020/147404 A1

- (51) 国际专利分类号:
G10L 13/02 (2013.01)
- (21) 国际申请号: PCT/CN2019/117775
- (22) 国际申请日: 2019 年 11 月 13 日 (13.11.2019)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
201910042827.1 2019 年 1 月 17 日 (17.01.2019) CN
- (71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
- (72) 发明人: 陈闽川(CHEN, Minchuan); 中国广东省深圳市福田区福田街道福安社区益田路

5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。马骏(MA, Jun); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。王少军(WANG, Shaojun); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 北京汇思诚业知识产权代理有限公司(UNI-INTEL PATENT AND TRADEMARK LAW FIRM); 中国北京市朝阳区建国门外大街永安东里甲 3 号通用国际中心 A 座 3 层, Beijing 100022 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS,

(54) **Title:** TEXT-TO-SPEECH SYNTHESIS METHOD, DEVICE, COMPUTER APPARATUS, AND NON-VOLATILE COMPUTER READABLE STORAGE MEDIUM

(54) 发明名称: 文本的语音合成方法、装置、计算机设备及计算机非易失性可读存储介质

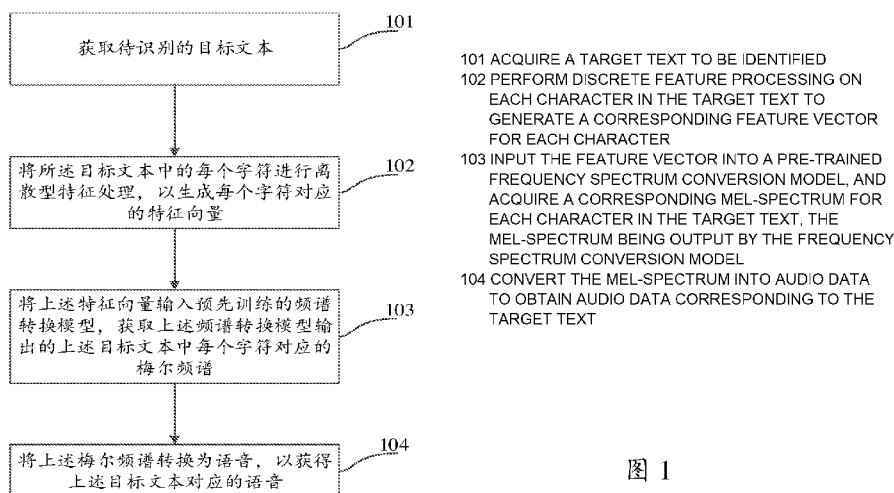


图 1

(57) **Abstract:** A text-to-speech synthesis method, a device, and a computer apparatus. The text-to-speech synthesis method comprises: first acquiring a target text to be identified (101); performing discrete feature processing on each character in the target text to generate a corresponding feature vector for each character (102); inputting the feature vector into a pre-trained frequency spectrum conversion model, and acquiring a corresponding Mel-spectrum for each character in the target text, the Mel-spectrum being output by the frequency spectrum conversion model (103); and converting the Mel-spectrum into audio data to obtain audio data corresponding to the target text (104). Thus, speech synthesis is performed without generating phonemic notation of each character in a text so as to effectively reduce a workload during a speech synthesis process, provide an effective solution for pronunciation issues during the speech synthesis process, and achieve a wide application range in the field of artificial intelligence.



JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW。

(84) 指定国(除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

— 包括国际检索报告(条约第21条(3))。

(57) 摘要: 一种文本的语音合成方法、装置和计算机设备, 其中, 上述文本的语音合成方法包括首先获取待识别的目标文本(101); 然后将所述目标文本中的每个字符进行离散型特征处理, 以生成每个字符对应的特征向量(102); 其次将所述特征向量输入预先训练的频谱转换模型, 获取上述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱(103); 最后再将所述梅尔频谱转换为语音, 以获得所述目标文本对应的语音(104)。这样, 在进行语音合成的时候, 无需对文本中的每个字符进行拼音标注, 有效地减少了语音合成过程中的工作量, 并为语音合成过程中的发音问题提供了有效的解决方案, 在人工智能领域中具有更广泛的应用范围。

文本的语音合成方法、装置、计算机设备及计算机非易失性可读存储介质

本申请要求于 2019 年 1 月 17 日提交中国专利局、申请号为 201910042827.1、申请名称为“文本的语音合成方法、装置和计算机设备”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

【技术领域】

本申请涉及人工智能技术领域，尤其涉及一种文本的语音合成方法、装置和计算机设备。

【背景技术】

由人工通过一定的机器设备产生出语音称为语音合成。语音合成是人机语音通信的一个重要组成部分。利用语音合成技术可以让机器像人一样说话，使一些以其他方式表示或存储的信息能转换为语音，从而人们可以通过听觉方便地获得这些信息。

现有相关技术中，为解决语音合成技术中的多音字发音问题，大多采用基于规则的方法或者基于统计机器学习的方法，但是，基于规则的方法需要人工设定大量的规则，基于统计机器学习的方法容易受到样本不均匀分布的限制。不仅如此，上述基于规则的方法和基于统计机器学习的方法均需要对训练文本进行大量的语音标注，这无疑大大增加了工作量。

【申请内容】

本申请实施例提供一种文本的语音合成方法、装置和计算机设备，通过预先训练的频谱转换模型，可以获取待识别文本中每个字符对应的梅尔频谱，并将梅尔频谱转换为语音，据此可以获得目标文本对应的语音，有效地减少了语音合成过程中的工作量，并为语音合成过程中的发音问题提供了有效的解决方案。

第一方面，本申请实施例提供了一种文本的语音合成方法，包括：

获取待识别的目标文本；

将所述目标文本中的每个字符进行离散型特征处理，以生成每个字符对应的特征向量；

将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱；

将所述梅尔频谱转换为语音，以获得所述目标文本对应的语音。

第二方面，本申请实施例还提供了一种文本的语音合成装置，包括：

获取模块，用于获取待识别的目标文本，并用于获取通过处理模块进行离散化特征处理之后的所述目标文本中的每个字符对应的特征向量，以及将所述目标文本中每个字符对应的特征向量输入至预先训练的频谱转换模型，以获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱；

转换模块，用于将所述获取模块获取的梅尔频谱转换为语音，以获得所述目标文本对应的语音。

第三方面，本申请实施例还提供了一种计算机设备，包括存储器、处理器及存储在该存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时，实现上述文本的语音合成方法。

第四方面，本申请实施例还提供了一种计算机非易失性可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现上述文本的语音合成方法。

以上技术方案中，获取待识别的目标文本之后，将所述目标文本中的每个字符进行离散型特征处理，以生成每个字符对应的特征向量，并将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱，以及将所述梅尔频谱转换为语音，以获得所述目标文本对应的语音，这样在进行语音合成的时候，便无需对文本中的每个字符进行拼音标注，有效地减少了语音合成过程中的工作量，并为语音合成过程中的发音问题提供了有效的解决方案。

【附图说明】

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有技术描述中所需要使用的附图作一简单地介绍，显而易见地，下面描述中的附图是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动性的前提下，还可以根据这些附图获得其他的附图。

图 1 为本申请文本的语音合成方法一个实施例的流程图；

图 2 为本申请文本的语音合成方法另一个实施例的流程图；

图 3 为本申请文本的语音合成装置一个实施例的连接结构示意图；

图 4 为本申请计算机设备一个实施例的结构示意图。

5 【具体实施方式】

为了更好的理解本申请的技术方案，下面结合附图对本申请实施例进行详细描述。

应当明确，所描述的实施例仅仅是本申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例，都属于本申请保护的范围。

在本申请实施例中使用的术语是仅仅出于描述特定实施例的目的，而非旨在限制本申请。在本申请实施例和所附权利要求书中所使用的单数形式的“一种”、“所述”和“该”也旨在包括多数形式，除非上下文清楚地表示其他含义。

图 1 为本申请文本的语音合成方法一个实施例的流程图，如图 1 所示，上述方法可以包括：

步骤 101：获取待识别的目标文本。

具体来说，可通过获取模块获取上述待识别的文本，上述获取模块可以是任意具有书面语言表现功能的输入法，其中，上述目标文本指的是任意具有书面语言表现形式的一段文字。

步骤 102：将所述目标文本中的每个字符进行离散型特征处理，以生成每个字符对应的特征向量。

进一步而言，离散型特征处理主要用于将连续的数值属性转化为离散的数值属性。本申请采取独热（One-Hot）编码的方式进行目标文本的离散型特征处理。

具体地，以下将对本申请如何使用 One-Hot 编码得到上述目标文本中每个字符对应的特征向量进行说明：

首先，假设本申请具有如下预设关键词，且每个关键词的编号如下所示：

1 老师；2 喜欢；3 学识；4 上课；5 很；6 幽默；7 我；8 渊博。

其次，当本申请中的目标文本为“老师学识很渊博”时，首先对该目标文本进行分离以与上述预设关键词相匹配，即分离成“老师”、“学识”、“很”以及“渊博”。

紧接着，将上述“老师”、“学识”、“很”以及“渊博”与上述预设关键词的编号一一对应，即得到：

1 老师	2 喜欢	3 学识	4 上课	5 很	6 幽默	7 我	8 渊博
1	0	1	0	1	0	0	1

因此，对于目标文本“老师学识很渊博”，最终可得到该目标文本中每个字符对应的特征向量为 10101001。

- 5 其中，上述预设关键词以及预设关键词的编号可以在具体实现时，根据实现需求进行自行设定，本实施例对上述预设关键词以及预设关键词对应的编号不作限定，以上预设关键词以及预设关键词的编号为一示例以便理解。

步骤 103：将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱。

- 10 在具体实现时，上述频谱转换模型可以为序列转换模型（Sequence to Sequence；以下简称：seq2seq）。进一步来说，本申请通过 seq2seq 模型输出与目标文本中的每个字符对应的梅尔频谱，由于 seq2seq 模型是目前自然语言处理技术中非常重要而且非常流行的一个模型，因此有着良好的表现。其中，本申请采取梅尔频谱作为声音特征的表现形式，可以使得人耳可以更容易感知到声音频率的变化。

- 具体而言，声音频率的单位是赫兹，人耳能听到的频率范围是 20 至 20000Hz，但人耳对赫兹这种标度单位并不是线性感知关系。例如我们适应了 1000Hz 的音调，如果把音调频率提高到 2000Hz，我们的耳朵只能觉察到频率提高了一点点，根本察觉不到频率提高了一倍。而通过梅尔频谱的表现方式，则人耳对频率的感知度就成了线性关系。也就是说，如果两端语音的梅尔频率相差两倍，则人耳可以感知到的音调大概也相差两倍。

步骤 104：将上述梅尔频谱转换为语音，以获得所述目标文本对应的语音。

- 25 进一步而言，可通过在频谱转换模型的外部接一声码器，以将上述梅尔频谱转换为语音实现输出。

- 在实际应用中，上述声码器可通过傅里叶逆变换将上述梅尔频谱转换为时域的语音波形信号，由于时域是真实世界，是唯一实际存在的域，故本申请可更为形象与直观的获取上述语音。上述语音合成方法中，获取待识别的目标文本之后，将所述目标文本中的每个字符进行离散型特征处理，以生成每个字符对应的特征向量，并将所述特征向量输入预先训练的频谱转换模型，

获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱，以及将所述梅尔频谱转换为语音，以获得所述目标文本对应的语音，这样在进行语音合成的时候，便无需对文本中的每个字符进行拼音标注，有效地减少了语音合成过程中的工作量，并为语音合成过程中的发音问题提供了有效的解决方案。

图2为本申请文本的语音合成方法另一个实施例的流程图，如图2所示，本申请图1所示实施例中，步骤103之前，还可以包括：

步骤201：获取预设数量的训练文本和上述训练文本对应的配对语音。

具体来说，和目标文本的概念类似，本实施例中的训练文本也指任意具有书面语言表现形式的一段文字。

其中，上述预设数量可以在具体实现时，根据系统性能和/或实现需求等自行设定，本实施例对上述预设数量的大小不作限定，举例来说，上述预设数量可以是1000个。

步骤202：对上述训练文本进行离散型特征处理，以获得上述训练文本中的每个字符对应的特征向量。

同样，在本实施例中，可以采用One-Hot编码的方式进行训练文本的离散型特征处理。具体实现过程可参考步骤102中的相关描述，在此不再赘述。

步骤203：将上述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得上述待训练的频谱转换模型输出的梅尔频谱。

进一步来说，上述步骤203可以包括：

步骤（1）通过上述待训练的频谱转换模型，对上述训练文本进行编码，获得与上述训练文本对应的隐藏状态序列，上述隐藏状态序列包括至少两个隐节点。

其中，上述隐藏状态序列的获得为将上述训练文本中每个字符的特征向量一一进行映射，其中上述训练文本中的字符的个数与隐节点的个数一一对应。

步骤（2）根据每个字符对应的隐节点的权值，分别对上述隐节点进行加权计算，获得上述训练样本中的每个字符对应的语义向量。

具体来说，可通过采取注意力机制的公式（1）来获得对应的语义向量：

$$C_i = \sum_{j=1}^N a_{ij} h_j \quad (1)$$

其中， C_i 表示第*i*个语义向量，*N*表示隐节点的个数， h_j 表示编码时的第

j 个字符的隐节点, 上述注意力机制指的是: 由于 a_{ij} 表示在编码时的第 j 阶段和解码时的第 i 阶段的相关性, 故对于每个语义向量均会选取与当前所要输出最合适的上下文信息。

步骤 (3) 对每个字符对应的语义向量进行解码, 并输出每个字符对应的梅尔频谱。

步骤 204: 当上述待训练的频谱转换模型输出的梅尔频谱与上述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时, 获得训练好的频谱转换模型。

进一步地, 当上述待训练的频谱转换模型输出的梅尔频谱与上述配对语音对应的梅尔频谱之间的误差大于预设阈值时, 还包括:

对每个隐节点的权值进行反向传播误差信息以进行更新, 不断迭代, 直至通过上述待训练的频谱转换模型输出的梅尔频谱与上述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值。

具体来说, 在将上述隐节点的权值进行更新之后, 首先需要对更新权值后的隐节点进行加权计算, 以获得训练样本中的每个字符对应的语义向量, 然后对每个字符对应的语义向量进行解码, 并输出每个字符对应的梅尔频谱, 最后当所述每个字符对应的梅尔频谱与上述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时, 停止上述每个隐节点的更新过程, 并获得训练好的频谱转换模型。

其中, 上述预设阈值可以在具体实现时, 根据系统性能和/或实现需求等自行设定, 本实施例对上述预设阈值的大小不作限定, 举例来说, 上述预设阈值可以为 80%。

图 3 为本申请文本的语音合成装置一个实施例的连接结构示意图, 如图 3 所示, 上述装置包括获取模块 31 和转换模块 32:

获取模块 31, 用于获取待识别的目标文本, 并用于获取通过处理模块 33 进行离散化特征处理之后的所述目标文本中的每个字符对应的特征向量, 以及将所述目标文本中每个字符对应的特征向量输入至预先训练的频谱转换模型, 以获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱。

具体来说, 可通过任意具有书面语言表现功能的输入法获取待识别的目标文本, 其中, 上述目标文本指的是任意具有书面语言表现形式的一段文字。

在具体实现时，上述频谱转换模型可以为序列转换模型（Sequence to Sequence；以下简称：seq2seq）。进一步来说，本申请通过 seq2seq 模型输出与目标文本中的每个字符对应的梅尔频谱，由于 seq2seq 模型是目前自然语言处理技术中非常重要而且非常流行的一个模型，因此有着良好的表现。其中，本申请采取梅尔频谱作为声音特征的表现形式，可以使得人耳可以更容易感知到声音频率的变化。

具体而言，声音频率的单位是赫兹，人耳能听到的频率范围是 20 至 20000Hz，但人耳对赫兹这种标度单位并不是线性感知关系。例如我们适应了 1000Hz 的音调，如果把音调频率提高到 2000Hz，我们的耳朵只能觉察到频率提高了一点点，根本察觉不到频率提高了一倍。而通过梅尔频谱的表现方式，则人耳对频率的感知度就成了线性关系。也就是说，如果两端语音的梅尔频率相差两倍，则人耳可以感知到的音调大概也相差两倍。

进一步而言，本申请采取独热（One-Hot）编码的方式进行目标文本的离散型特征处理。然后再将上述特征向量输入至预先训练的频谱转换模型，最终获取上述频谱转换模型输出的目标文本中每个字符对应的梅尔频谱。

进一步，以下将对本申请如何使用 One-Hot 编码得到上述目标文本中每个字符对应的特征向量进行说明：

首先，假设本申请具有如下预设关键词，且每个关键词的编号如下所示：

1 老师；2 喜欢；3 学识；4 上课；5 很；6 幽默；7 我；8 渊博。

其次，当本申请中的目标文本为“老师学识很渊博”时，首先对该目标文本进行分离以与上述预设关键词相匹配，即分离成“老师”、“学识”、“很”以及“渊博”。

紧接着，将上述“老师”、“学识”、“很”以及“渊博”与上述预设关键词的编号一一对应，即得到：

1 老师	2 喜欢	3 学识	4 上课	5 很	6 幽默	7 我	8 渊博
1	0	1	0	1	0	0	1

因此，对于目标文本“老师学识很渊博”，最终可得到该目标文本中每个字符对应的特征向量为 10101001。

其中，上述预设关键词以及预设关键词的编号可以在具体实现时，根据实现需求进行自行设定，本实施例对上述预设关键词以及预设关键词对应的编号不作限定，以上预设关键词以及预设关键词的编号为一示例以便理解。

转换模块 32，用于将上述获取模块 31 获取的梅尔频谱转换为语音，以获

得上述目标文本对应的语音。

进一步而言，上述转换模块 32 可以是一声码器，上述声码器在进行转换处理时，可通过傅里叶逆变换将上述梅尔频谱转换为时域的语音波形信号，由于时域是真实世界，是唯一实际存在的域，故本申请可更为形象与直观的获取上述语音。

上述语音合成装置中，获取模块 31 获取待识别的目标文本之后，通过处理模块 33 将上述目标文本中的每个字符进行离散型特征处理，以生成每个字符对应的特征向量，并将上述特征向量输入预先训练的频谱转换模型，获取上述频谱转换模型输出的上述目标文本中每个字符对应的梅尔频谱，以及通过转换模块 32 将上述梅尔频谱转换为语音，以获得上述目标文本对应的语音，这样在进行语音合成的时候，便无需对文本中的每个字符进行拼音标注，有效地减少了语音合成过程中的工作量，并为语音合成过程中的发音问题提供了有效的解决方案。

再请参阅图 3 所示，在另一实施例中，：

所述获取模块 31 还用于在将上述特征向量输入预先训练的频谱转换模型，获取上述频谱转换模型输出的上述目标文本中每个字符对应的梅尔频谱之前，获取预设数量的训练文本和上述训练文本对应的配对语音，并用于获取通过所述处理模块 33 对所述训练文本进行离散型特征处理之后的所述训练文本中的每个字符对应的特征向量，且将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱；以及，当所述待训练的频谱转换模型输出的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时，获得训练好的频谱转换模型。

具体来说，和目标文本的概念类似，本实施例中的训练文本也指任意具有书面语言表现形式的一段文字。

其中，上述预设数量可以在具体实现时，根据系统性能和/或实现需求等自行设定，本实施例对上述预设数量的大小不作限定，举例来说，上述预设数量可以是 1000 个。

同样，在本实施例中，通过处理模块 33 对上述训练文本进行离散型特征处理，以获得上述训练文本中的每个字符对应的特征向量，在具体实现时，可以采用 One-Hot 编码的方式进行训练文本的离散型特征处理。具体实现

过程可参考图 3 所示实施例的相关描述，在此不再赘述。

进一步来说，上述获取模块 31 获取的上述预设数量的配对语音对应的梅尔频谱可以包括：

步骤（1）通过上述待训练的频谱转换模型，对上述训练文本进行编码，
5 获得与上述训练文本对应的隐藏状态序列，上述隐藏状态序列包括至少两个隐节点。

其中，上述隐藏状态序列的获得为将上述训练文本中每个字符的特征向量一一进行映射，其中上述训练文本中的字符的个数与隐节点的个数一一对应。

10 步骤（2）根据每个字符对应的隐节点的权值，分别对上述隐节点进行加权计算，获得上述训练样本中的每个字符对应的语义向量。

具体来说，可通过采取注意力机制的公式（1）来获得对应的语义向量：

$$C_i = \sum_{j=1}^N a_{ij} h_j \quad (1)$$

15 其中， C_i 表示第 i 个语义向量， N 表示隐节点的个数， h_j 表示编码时的第 j 个字符的隐节点，上述注意力机制指的是：由于 a_{ij} 表示在编码时的第 j 阶段和解码时的第 i 阶段的相关性，故对于每个语义向量均会选取与当前所要输出最合适的上下文信息。

步骤（3）对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱。

20 上述获取模块 31 具体用于通过待训练的频谱转换模型，对上述训练文本进行编码，获得与上述训练文本对应的隐藏状态序列，上述隐藏状态序列包括至少两个隐节点，并根据每个字符对应的隐节点的权值，分别对上述隐节点进行加权计算，获得上述训练样本中的每个字符对应的语义向量，以及对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱。

25 进一步地，当上述待训练的频谱转换模型输出的梅尔频谱与上述配对语音对应的梅尔频谱之间的误差大于预设阈值时，还包括：

对每个隐节点的权值进行反向传播误差信息以进行更新，不断迭代，直至通过上述待训练的频谱转换模型输出的梅尔频谱与上述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值。

30 具体来说，在将上述隐节点的权值进行更新之后，首先需要对更新权值后的隐节点进行加权计算，以获得训练样本中的每个字符对应的语义向量，

然后对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱，最后当所述每个字符对应的梅尔频谱与上述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时，停止上述每个隐节点的更新过程，并获得训练好的频谱转换模型。

- 5 其中，上述预设阈值可以在具体实现时，根据系统性能和/或实现需求等自行设定，本实施例对上述预设阈值的大小不作限定，举例来说，上述预设阈值可以为 80%。

10 图 4 为本申请计算机设备一个实施例的结构示意图，上述计算机设备可以包括存储器、处理器及存储在上述存储器上并可在上述处理器上运行的计算机程序，上述处理器执行上述计算机程序时，可以实现本申请实施例提供的文本的语音合成方法。

15 其中，上述计算机设备可以为服务器，例如：云服务器，或者上述计算机设备也可以为电子设备，例如：智能手机、智能手表、个人计算机（Personal Computer；以下简称：PC）、笔记本电脑或平板电脑等智能设备，本实施例对上述计算机设备的具体形态不作限定。

图 4 示出了适于用来实现本申请实施方式的示例性计算机设备 52 的框图。图 4 显示的计算机设备 52 仅仅是一个示例，不对本申请实施例的功能和使用范围带来任何限制。

20 如图 4 所示，计算机设备 52 以通用计算设备的形式表现。计算机设备 52 的组件可以包括但不限于：一个或者多个处理器或者处理单元 56，系统存储器 78，连接不同系统组件（包括系统存储器 78 和处理单元 56）的总线 58。

25 总线 58 表示几类总线结构中的一种或多种，包括存储器总线或者存储器控制器，外围总线，图形加速端口，处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说，这些体系结构包括但不限于工业标准体系结构（Industry Standard Architecture；以下简称：ISA）总线，微通道体系结构（Micro Channel Architecture；以下简称：MAC）总线，增强型 ISA 总线、视频电子标准协会（Video Electronics Standards Association；以下简称：VESA）局域总线以及外围组件互连（Peripheral Component Interconnection；以下简称：30 PCI）总线。

计算机设备 52 典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算机设备 52 访问的可用介质，包括易失性和非易失性介质，可移

动的和不可移动的介质。

系统存储器 78 可以包括易失性存储器形式的计算机系统可读介质，例如随机存取存储器 (Random Access Memory; 以下简称: RAM) 70 和/或高速缓存存储器 72。计算机设备 52 可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例，存储系统 74 可以用于读写不可移动的、非易失性磁介质 (图 4 未显示，通常称为“硬盘驱动器”)。尽管图 4 中未示出，可以提供用于对可移动非易失性磁盘 (例如“软盘”) 读写的磁盘驱动器，以及对可移动非易失性光盘 (例如: 光盘只读存储器 (Compact Disc Read Only Memory; 以下简称: CD-ROM)、数字多功能只读光盘 (Digital Video Disc Read Only Memory; 以下简称: DVD-ROM) 或者其它光介质) 读写的光盘驱动器。在这些情况下，每个驱动器可以通过一个或者多个数据介质接口与总线 58 相连。存储器 78 可以包括至少一个程序产品，该程序产品具有一组 (例如至少一个) 程序模块，这些程序模块被配置以执行本申请各实施例的功能。

具有一组 (至少一个) 程序模块 82 的程序/实用工具 80，可以存储在例如存储器 78 中，这样的程序模块 82 包括——但不限于——操作系统、一个或者多个应用程序、其它程序模块以及程序数据，这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块 82 通常执行本申请所描述的实施例中的功能和/或方法。

计算机设备 52 也可以与一个或多个外部设备 54 (例如键盘、指向设备、显示器 64 等) 通信，还可与一个或者多个使得用户能与该计算机设备 52 交互的设备通信，和/或与使得该计算机设备 52 能与一个或多个其它计算设备进行通信的任何设备 (例如网卡，调制解调器等等) 通信。这种通信可以通过输入/输出 (I/O) 接口 62 进行。并且，计算机设备 52 还可以通过网络适配器 60 与一个或者多个网络 (例如局域网 (Local Area Network; 以下简称: LAN)，广域网 (Wide Area Network; 以下简称: WAN) 和/或公共网络，例如因特网) 通信。如图 4 所示，网络适配器 60 通过总线 58 与计算机设备 52 的其它模块通信。应当明白，尽管图 4 中未示出，可以结合计算机设备 52 使用其它硬件和/或软件模块，包括但不限于: 微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、RAID 系统、磁带驱动器以及数据备份存储系统等。

处理单元 56 通过运行存储在系统存储器 78 中的程序，从而执行各种功能应用以及数据处理，例如实现本申请实施例提供的文本的语音合成方

法。

本申请实施例还提供一种计算机非易失性可读存储介质，其上存储有计算机程序，上述计算机程序被处理器执行时可以实现本申请实施例提供的文本的语音合成方法。

5 上述计算机非易失性可读存储介质可以采用一个或多个计算机可读的介质的任意组合。计算机可读介质可以是计算机可读信号介质或者计算机可读存储介质。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器（RAM）、只读存储器（Read
10 Only Memory；以下简称：ROM）、可擦式可编程只读存储器（Erasable Programmable Read Only Memory；以下简称：EPROM）或闪存、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储
15 程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

计算机可读的信号介质可以包括在基带中或者作为载波一部分传播的数据信号，其中承载了计算机可读的程序代码。这种传播的数据信号可以采用多种形式，包括——但不限于——电磁信号、光信号或上述的任意合适的组合。
20 计算机可读的信号介质还可以是计算机可读存储介质以外的任何计算机可读介质，该计算机可读介质可以发送、传播或者传输用于由指令执行系统、装置或者器件使用或者与其结合使用的程序。

计算机可读介质上包含的程序代码可以用任何适当的介质传输，包括——但不限于——无线、电线、光缆、RF 等等，或者上述的任意合适的组合。

25 可以以一种或多种程序设计语言或其组合来编写用于执行本申请操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言——诸如 Java、Smalltalk、C++，还包括常规的过程式程序设计语言——诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在
30 远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网（Local Area Network；以下简称：LAN）或广域网（Wide Area Network；以下简称：

WAN) 连接到用户计算机, 或者, 可以连接到外部计算机 (例如利用因特网服务提供商来通过因特网连接)。

在本说明书的描述中, 参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中, 对上述术语的示意性表述不必须针对的是相同的实施例或示例。而且, 描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外, 在不相互矛盾的情况下, 本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

此外, 术语“第一”、“第二”仅用于描述目的, 而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此, 限定有“第一”、“第二”的特征可以明示或者隐含地包括至少一个该特征。在本申请的描述中, “多个”的含义是至少两个, 例如两个, 三个等, 除非另有明确具体的限定。

流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为, 表示包括一个或更多个用于实现定制逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分, 并且本申请的优选实施方式的范围包括另外的实现, 其中可以不按所示出或讨论的顺序, 包括根据所涉及的功能按基本同时的方式或按相反的顺序, 来执行功能, 这应被本申请的实施例所属技术领域的技术人员所理解。

取决于语境, 如在此所使用的词语“如果”可以被解释成为“在……时”或“当……时”或“响应于确定”或“响应于检测”。类似地, 取决于语境, 短语“如果确定”或“如果检测 (陈述的条件或事件)”可以被解释成为“当确定时”或“响应于确定”或“当检测 (陈述的条件或事件) 时”或“响应于检测 (陈述的条件或事件)”。

需要说明的是, 本申请实施例中所涉及的终端可以包括但不限于个人计算机 (Personal Computer; 以下简称: PC)、个人数字助理 (Personal Digital Assistant; 以下简称: PDA)、无线手持设备、平板电脑 (Tablet Computer)、手机、MP3 播放器、MP4 播放器等。

在本申请所提供的几个实施例中, 应该理解到, 所揭露的系统、装置和方法, 可以通过其它的方式实现。例如, 以上所描述的装置实施例仅仅是示意性的, 例如, 所述单元的划分, 仅仅为一种逻辑功能划分, 实际实现时可以

有另外的划分方式，例如，多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

5 另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能单元的形式实现。

10 上述以软件功能单元的形式实现的集成的单元，可以存储在一个计算机可读取存储介质中。上述软件功能单元存储在一个存储介质中，包括若干指令用以使得一台计算机装置（可以是个人计算机，服务器，或者网络装置等）或处理器（Processor）执行本申请各个实施例所述方法的部分步骤。而前述的存储介质包括：U 盘、移动硬盘、只读存储器（Read-Only Memory；以下简称：ROM）、随机存取存储器（Random Access Memory；以下简称：RAM）、
15 磁碟或者光盘等各种可以存储程序代码的介质。

 以上所述仅为本申请的较佳实施例而已，并不用以限制本申请，凡在本申请的精神和原则之内，所做的任何修改、等同替换、改进等，均应包含在本申请保护的范围之内。

权 利 要 求 书

1、一种文本的语音合成方法，其特征在于，所述的方法包括：

获取待识别的目标文本；

将所述目标文本中的每个字符进行离散型特征处理，以生成每个字符对应的特征向量；

将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱；

将所述梅尔频谱转换为语音，以获得所述目标文本对应的语音。

2、根据权利要求 1 所述的方法，其特征在于，所述将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱之前，还包括：

获取预设数量的训练文本和所述训练文本对应的配对语音；

对所述训练文本进行离散型特征处理，以获得所述训练文本中的每个字符对应的特征向量；

将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱；以及，

当所述待训练的频谱转换模型输出的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时，获得训练好的频谱转换模型。

3、根据权利要求 2 所述的方法，其特征在于，所述将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱包括：

通过所述待训练的频谱转换模型，对所述训练文本进行编码，获得与所述训练文本对应的隐藏状态序列，所述隐藏状态序列包括至少两个隐节点；

根据每个字符对应的隐节点的权值，分别对所述隐节点进行加权计算，获得所述训练样本中的每个字符对应的语义向量；

对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱。

4、根据权利要求 2 所述的方法，其特征在于，所述将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱之后，还包括：

当所述待训练的频谱转换模型输出的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差大于所述预设阈值时，对每个隐节点的权值进行更新；

分别对更新权值后的隐节点进行加权计算，以获得所述训练样本中的每个字符对应的语义向量；

对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱；以及，当所述每个字符对应的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时，停止对每个隐节点的权值的更新过程，并获得训练好的频谱转换模型。

5、根据权利要求 1-4 任意一项所述的方法，其特征在于，所述将所述梅尔频谱转换为语音，以获得所述目标文本对应的语音包括：

通过声码器对所述梅尔频谱进行傅里叶逆变换，以将所述梅尔频谱转换为时域的语音波形信号，获得所述语音。

6、一种文本的语音合成装置，其特征在于，所述的装置包括：

获取模块，用于获取待识别的目标文本，并用于获取通过处理模块进行离散化特征处理之后的所述目标文本中的每个字符对应的特征向量，以及将所述目标文本中每个字符对应的特征向量输入至预先训练的频谱转换模型，以获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱，所述处理模块用于对所述目标文本进行离散化特征处理以得到所述目标文本中的每个所述字符对应的所述特征向量；

转换模块，用于将所述获取模块获取的梅尔频谱转换为语音，以获得所述目标文本对应的语音。

7、根据权利要求 6 所述的装置，其特征在于，所述获取模块，还用于在将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱之前，获取预设数量的训练文本和所述训练文本对应的配对语音；对所述训练文本进行离散型特征处理，以获得所述训练文本中的每个字符对应的特征向量；将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱；以及，当所述待训练的频谱转换模型输出的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时，获得训练好的频谱转换模型。

8、根据权利要求 7 所述的装置，其特征在于，所述获取模块，具体用于通过待训练的频谱转换模型，对所述训练文本进行编码，获得与所述训练文本对应的隐藏状态序列，所述隐藏状态序列包括至少两个隐节点，并根据每个字符对应的隐节点的权值，分别对所述隐节点进行加权计算，获得所述训练样本中的每个字符对应的语义向量，以及对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱。

9、根据权利要求7所述的装置，其特征在于，所述获取模块，还用于当所述待训练的频谱转换模型输出的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差大于所述预设阈值时，对每个隐节点的权值进行更新；分别对更新权值后的隐节点进行加权计算，以获得所述训练样本中的每个字符对应的语义向量；对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱；以及，当所述每个字符对应的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时，停止对每个隐节点的权值的更新过程，并获得训练好的频谱转换模型。

10、根据权利要求6-9任意一项所述的装置，其特征在于，所述转换模块具体用于通过声码器对所述梅尔频谱进行傅里叶逆变换，以将所述梅尔频谱转换为时域的语音波形信号，获得所述语音。

11、一种计算机设备，其特征在于，包括存储器、处理器及存储在所述存储器上并可在所述处理器上运行的计算机程序，所述处理器执行所述计算机程序时，实现以下步骤：

获取待识别的目标文本；

将所述目标文本中的每个字符进行离散型特征处理，以生成每个字符对应的特征向量；

将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱；

将所述梅尔频谱转换为语音，以获得所述目标文本对应的语音。

12、根据权利要求11所述的计算机设备，其特征在于，在所述将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱之前，所述处理器还执行所述计算机程序实现以下步骤：

获取预设数量的训练文本和所述训练文本对应的配对语音；

对所述训练文本进行离散型特征处理，以获得所述训练文本中的每个字符对应的特征向量；

将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱；以及，

当所述待训练的频谱转换模型输出的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时，获得训练好的频谱转换模型。

13、根据权利要求 12 所述的计算机设备，其特征在于，所述处理器执行所述计算机程序以实现所述将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱时，包括以下步骤：

通过所述待训练的频谱转换模型，对所述训练文本进行编码，获得与所述训练文本对应的隐藏状态序列，所述隐藏状态序列包括至少两个隐节点；

根据每个字符对应的隐节点的权值，分别对所述隐节点进行加权计算，获得所述训练样本中的每个字符对应的语义向量；

对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱。

14、根据权利要求 12 所述的计算机设备，其特征在于，在所述将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱之后，所述处理器还执行所述计算机程序实现以下步骤：

当所述待训练的频谱转换模型输出的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差大于所述预设阈值时，对每个隐节点的权值进行更新；

分别对更新权值后的隐节点进行加权计算，以获得所述训练样本中的每个字符对应的语义向量；

对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱；以及，当所述每个字符对应的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时，停止对每个隐节点的权值的更新过程，并获得训练好的频谱转换模型。

15、根据权利要求 11-14 中任意一项所述的计算机设备，其特征在于，所述处理器执行所述计算机程序实现所述将所述梅尔频谱转换为语音以获得所述目标文本对应的语音时，包括以下步骤：

通过声码器对所述梅尔频谱进行傅里叶逆变换，以将所述梅尔频谱转换为时域的语音波形信号，获得所述语音。

16、一种计算机非易失性可读存储介质，其上存储有计算机程序，其特征在于，所述计算机程序被处理器执行时实现以下步骤：

获取待识别的目标文本；

将所述目标文本中的每个字符进行离散型特征处理，以生成每个字符对应的特征向量；

将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出

的所述目标文本中每个字符对应的梅尔频谱；

将所述梅尔频谱转换为语音，以获得所述目标文本对应的语音。

17、根据权利要求 16 所述的计算机非易失性可读存储介质，其特征在于，在所述将所述特征向量输入预先训练的频谱转换模型，获取所述频谱转换模型输出的所述目标文本中每个字符对应的梅尔频谱之前，所述计算机程序还被处理器执行实现以下步骤：

获取预设数量的训练文本和所述训练文本对应的配对语音；

对所述训练文本进行离散型特征处理，以获得所述训练文本中的每个字符对应的特征向量；

将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱；以及，

当所述待训练的频谱转换模型输出的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差小于或等于预设阈值时，获得训练好的频谱转换模型。

18、根据权利要求 17 所述的计算机非易失性可读存储介质，其特征在于，所述计算机程序被处理器执行实现所述将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱时，包括以下步骤：

通过所述待训练的频谱转换模型，对所述训练文本进行编码，获得与所述训练文本对应的隐藏状态序列，所述隐藏状态序列包括至少两个隐节点；

根据每个字符对应的隐节点的权值，分别对所述隐节点进行加权计算，获得所述训练样本中的每个字符对应的语义向量；

对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱。

19、根据权利要求 17 所述的计算机非易失性可读存储介质，其特征在于，在所述将所述训练文本中的每个字符对应的特征向量输入待训练的频谱转换模型，获得所述待训练的频谱转换模型输出的梅尔频谱之后，所述计算机程序还被处理器执行实现以下步骤：

当所述待训练的频谱转换模型输出的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差大于所述预设阈值时，对每个隐节点的权值进行更新；

分别对更新权值后的隐节点进行加权计算，以获得所述训练样本中的每个字符对应的语义向量；

对每个字符对应的语义向量进行解码，并输出每个字符对应的梅尔频谱；以及，当所述每个字符对应的梅尔频谱与所述配对语音对应的梅尔频谱之间的误差

小于或等于预设阈值时，停止对每个隐节点的权值的更新过程，并获得训练好的频谱转换模型。

20、根据权利要求 16-19 任意一项所述的计算机非易失性可读存储介质，其特征在于，所述计算机程序被处理器执行实现所述将所述梅尔频谱转换为语音以获得所述目标文本对应的语音时，包括以下步骤：

通过声码器对所述梅尔频谱进行傅里叶逆变换，以将所述梅尔频谱转换为时域的语音波形信号，获得所述语音。

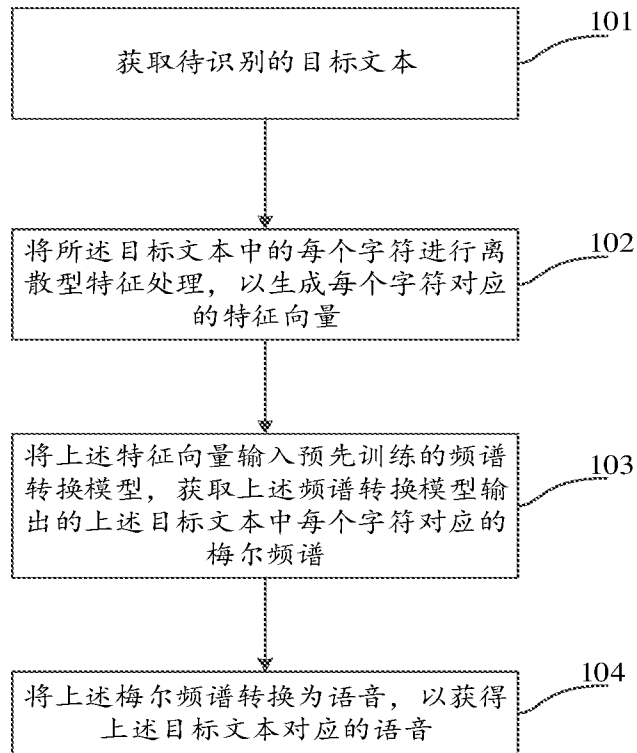


图 1

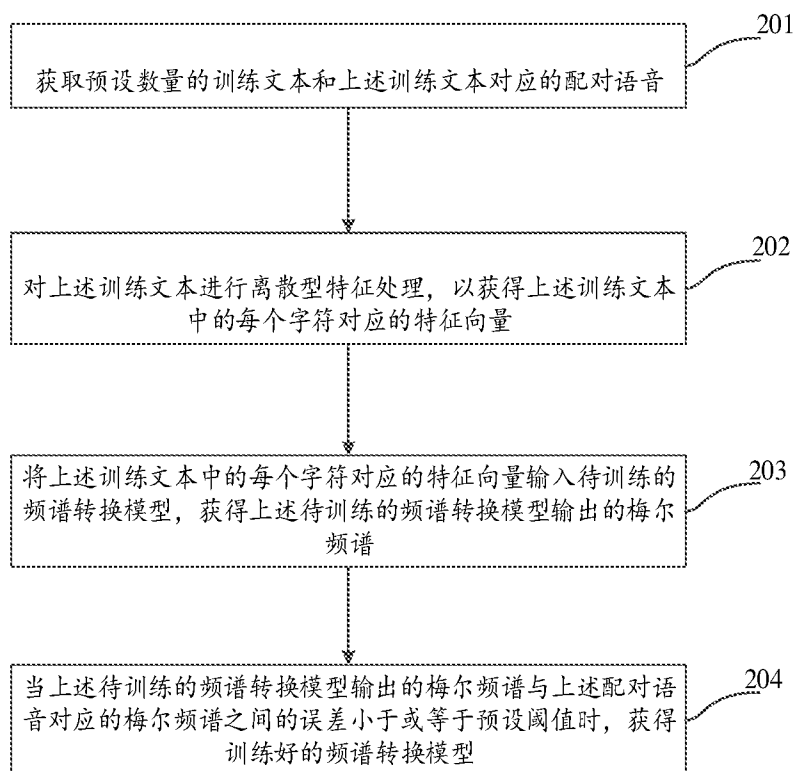


图 2

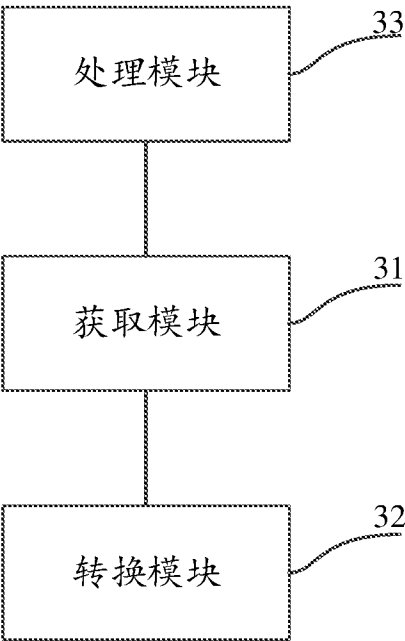


图 3

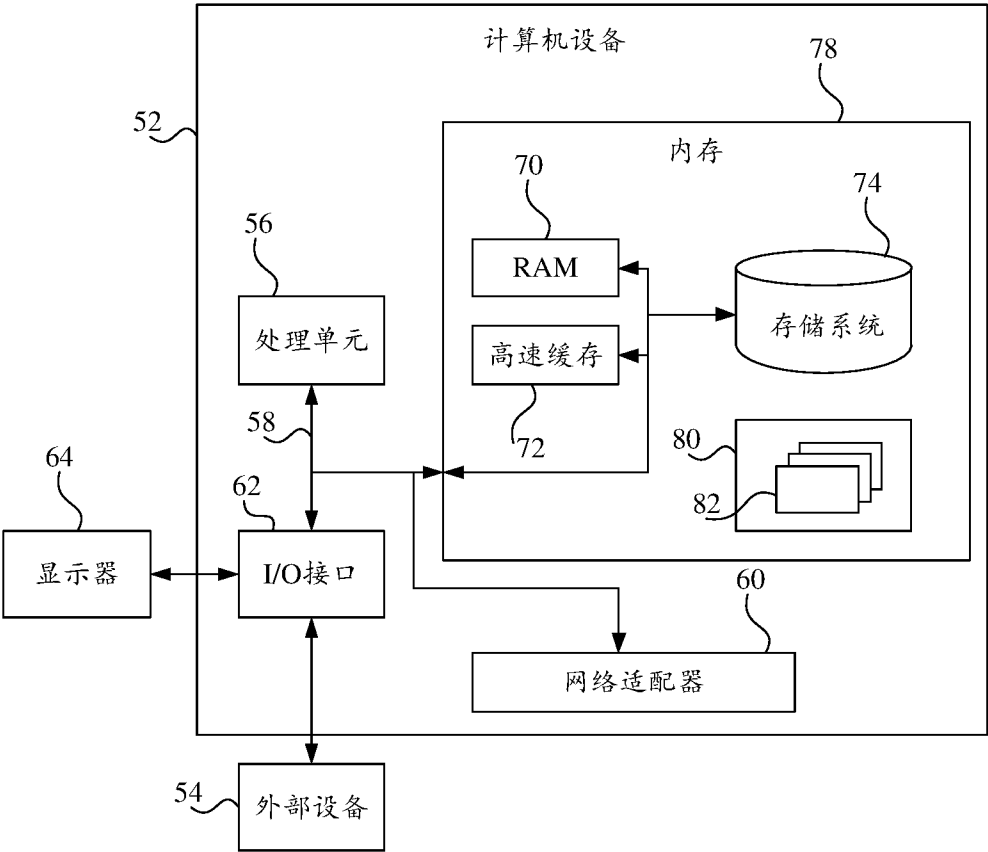


图 4

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/117775

A. CLASSIFICATION OF SUBJECT MATTER

G10L 13/02(2013.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G10L, G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNKI, CNPAT, WPI, EPODOC: 合成, 文本, 语音, 训练, 频谱, 特征, 向量, 模型, synthesis, text, voice, train, frequency spectrum, feature, vector, model

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 109754778 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 14 May 2019 (2019-05-14) description, paragraphs [0004]-[0032]	1-20
A	CN 107564511 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 09 January 2018 (2018-01-09) description, paragraphs [0003]-[0024]	1-20
A	CN 105654939 A (BEIJING TIMES RUILANG TECHNOLOGY CO., LTD.) 08 June 2016 (2016-06-08) entire document	1-10
A	US 2005119891 A1 (MICROSOFT CORPORATION) 02 June 2005 (2005-06-02) entire document	1-20
A	CN 109036375 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 18 December 2018 (2018-12-18) entire document	1-20
A	CN 108492818 A (BAIDU ONLINE NETWORK TECHNOLOGY (BEIJING) CO., LTD.) 04 September 2018 (2018-09-04) entire document	1-20

☐ Further documents are listed in the continuation of Box C.
☒ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

15 January 2020

Date of mailing of the international search report

23 January 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/117775

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	109754778	A	14 May 2019	None			
CN	107564511	A	09 January 2018	WO	2019056500	A1	28 March 2019
				HK	1246961	A0	14 September 2018
CN	105654939	A	08 June 2016	None			
US	2005119891	A1	02 June 2005	US	2002099547	A1	25 July 2002
				EP	1213705	A2	12 June 2002
				US	2004148171	A1	29 July 2004
				DE	60126564	D1	29 March 2007
				AT	354155	T	15 March 2007
				US	2002095289	A1	18 July 2002
CN	109036375	A	18 December 2018	None			
CN	108492818	A	04 September 2018	None			

国际检索报告

国际申请号

PCT/CN2019/117775

A. 主题的分类 G10L 13/02 (2013.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																							
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G10L, G06F 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNKI, CNPAT, WPI, EP0D0C: 合成, 文本, 语音, 训练, 频谱, 特征, 向量, 模型, synthesis, text, voice, train, frequency spectrum, feature, vector, model																							
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 109754778 A (平安科技深圳有限公司) 2019年 5月 14日 (2019 - 05 - 14) 说明书第4-32段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 107564511 A (平安科技深圳有限公司) 2018年 1月 9日 (2018 - 01 - 09) 说明书第3-24段</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 105654939 A (北京时代瑞朗科技有限公司) 2016年 6月 8日 (2016 - 06 - 08) 全文</td> <td>1-10</td> </tr> <tr> <td>A</td> <td>US 2005119891 A1 (MICROSOFT CORPORATION) 2005年 6月 2日 (2005 - 06 - 02) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 109036375 A (腾讯科技深圳有限公司) 2018年 12月 18日 (2018 - 12 - 18) 全文</td> <td>1-20</td> </tr> <tr> <td>A</td> <td>CN 108492818 A (百度在线网络技术北京有限公司) 2018年 9月 4日 (2018 - 09 - 04) 全文</td> <td>1-20</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	PX	CN 109754778 A (平安科技深圳有限公司) 2019年 5月 14日 (2019 - 05 - 14) 说明书第4-32段	1-20	A	CN 107564511 A (平安科技深圳有限公司) 2018年 1月 9日 (2018 - 01 - 09) 说明书第3-24段	1-20	A	CN 105654939 A (北京时代瑞朗科技有限公司) 2016年 6月 8日 (2016 - 06 - 08) 全文	1-10	A	US 2005119891 A1 (MICROSOFT CORPORATION) 2005年 6月 2日 (2005 - 06 - 02) 全文	1-20	A	CN 109036375 A (腾讯科技深圳有限公司) 2018年 12月 18日 (2018 - 12 - 18) 全文	1-20	A	CN 108492818 A (百度在线网络技术北京有限公司) 2018年 9月 4日 (2018 - 09 - 04) 全文	1-20
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 109754778 A (平安科技深圳有限公司) 2019年 5月 14日 (2019 - 05 - 14) 说明书第4-32段	1-20																					
A	CN 107564511 A (平安科技深圳有限公司) 2018年 1月 9日 (2018 - 01 - 09) 说明书第3-24段	1-20																					
A	CN 105654939 A (北京时代瑞朗科技有限公司) 2016年 6月 8日 (2016 - 06 - 08) 全文	1-10																					
A	US 2005119891 A1 (MICROSOFT CORPORATION) 2005年 6月 2日 (2005 - 06 - 02) 全文	1-20																					
A	CN 109036375 A (腾讯科技深圳有限公司) 2018年 12月 18日 (2018 - 12 - 18) 全文	1-20																					
A	CN 108492818 A (百度在线网络技术北京有限公司) 2018年 9月 4日 (2018 - 09 - 04) 全文	1-20																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
<table border="0"> <tr> <td> * 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 </td> <td> “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件 </td> </tr> </table>			* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																			
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件	“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件																						
国际检索实际完成的日期 2020年 1月 15日		国际检索报告邮寄日期 2020年 1月 23日																					
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 赵晓敏 电话号码 86-(10)-53961332																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/117775

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	109754778	A	2019年 5月 14日	无			
CN	107564511	A	2018年 1月 9日	WO	2019056500	A1	2019年 3月 28日
				HK	1246961	A0	2018年 9月 14日
CN	105654939	A	2016年 6月 8日	无			
US	2005119891	A1	2005年 6月 2日	US	2002099547	A1	2002年 7月 25日
				EP	1213705	A2	2002年 6月 12日
				US	2004148171	A1	2004年 7月 29日
				DE	60126564	D1	2007年 3月 29日
				AT	354155	T	2007年 3月 15日
				US	2002095289	A1	2002年 7月 18日
CN	109036375	A	2018年 12月 18日	无			
CN	108492818	A	2018年 9月 4日	无			