

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。