

(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



(10) 国际公布号
WO 2020/147404 A1

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

(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

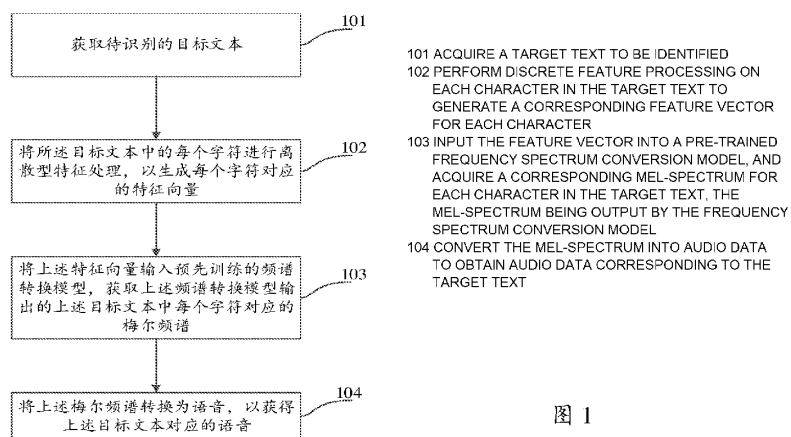


图 1

(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.

[见续页]

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。