(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
2026 年 8 月 27 日 (27.08.2026)

WIPO | PCT

(10) 国际公布号
WO 2026/174673 A1

(51) 国际专利分类号:
G10L 13/02 (2013.01)

(21) 国际申请号: PCT/CN2025/095832

(22) 国际申请日: 2025 年 5 月 19 日 (19.05.2025)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202510196407.4 2025年2月21日 (21.02.2025) CN

(71) 申请人: 无限星途科技私人有限公司 (**X STAR TECHNOLOGY PTE. LTD.**) [SG/SG]; 新加坡新达城1号楼, 37-03, 淡马锡林荫道7号 038987 (SG)。

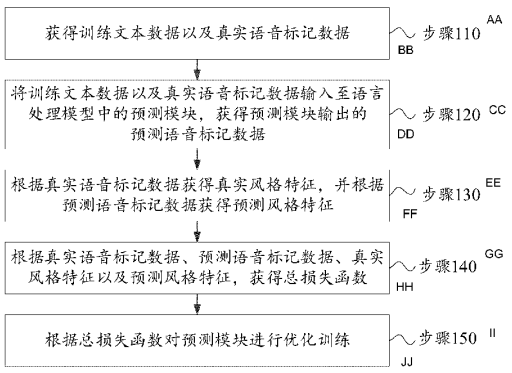
(72) 发明人: 张宇强 (**ZHANG, Yuqiang**); 中国上海市浦东新区中国（上海）自由贸易试验区浦东大道2123号三层 200135 (CN)。 王明超 (**WANG, Mingchao**); 中国上海市浦东新区中国（上海）自由贸易试验区浦东大道2123号三层 200135 (CN)。 张磊 (**ZHANG, Lei**); 中国上海市浦东新区中国（上海）自由贸易试验区浦东大道2123号三层 200135 (CN)。 刘兵 (**LIU, Bing**); 中国上海市浦东新区中国（上海）自由贸易试验区浦东大道2123号三层 200135 (CN)。 胡国星 (**HU, Guoxing**); 中国上海市浦东新区中国（上海）自由贸易试验区浦东大道2123号三层 200135 (CN)。 唐守旭 (**TANG, Shouxu**); 中国上海市浦东新区中国（上海）自由贸易试验区浦东大道2123号三层 200135 (CN)。

(74) 代理人: 北京超凡宏宇知识产权代理有限公司 (**CHOFN INTELLECTUAL PROPERTY**); 中国北京市海淀区北四环西路68号左岸工社1215-1218室 100080 (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,

(54) **Title:** LANGUAGE PROCESSING MODEL TRAINING METHOD AND APPARATUS, ELECTRONIC DEVICE, AND READABLE STORAGE MEDIUM

(54) 发明名称: 语言处理模型训练方法、装置、电子设备和可读存储介质



AA Step 110
BB Obtain training text data and real speech labeling data
CC Step 120
DD Input the training text data and the real speech labeling data into a prediction module in a language processing model to obtain predicted speech labeling data outputted by the prediction module
EE Step 130
FF Obtain a real style feature on the basis of the real speech labeling data, and obtain a predicted style feature on the basis of the predicted speech labeling data
GG Step 140
HH Obtain a total loss function on the basis of the real speech labeling data, the predicted speech labeling data, the real style feature, and the predicted style feature
II Step 150
JJ Perform an optimization training on the prediction module on the basis of the total loss function

(57) **Abstract:** A language processing model training method and apparatus, an electronic device, and a readable storage medium, relating to the technical field of speech synthesis. The method comprises: obtaining training text data and real speech labeling data (step 110); inputting the training text data and the real speech labeling data into a prediction module in a language processing model to obtain predicted speech labeling data outputted by the prediction module (step 120); obtaining a real style feature on the basis of the real speech labeling data, and obtaining a predicted style feature on the basis of the predicted speech labeling data (step 130); obtaining a total loss function on the basis of the real speech labeling data, the predicted speech labeling data, the real style feature, and the predicted style feature (step 140); and performing an optimization training on the prediction module on the basis of the total loss function (step 150). The style feature loss is introduced into a total loss function, so that it is ensured that a language processing model can learn about style information, and thus the language processing model reproduces a speech style more accurately, thereby improving the performance of the language processing model in processing speech style information. Additionally, a speech processing model is a key part in a speech synthesis model, thereby improving the performance of the large speech synthesis model in processing speech style information.

WO 2026/174673 A1

[见续页]

CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, 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, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UY, UZ, VC, VN, WS, 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, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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: 语言处理模型训练方法、装置、电子设备和可读存储介质，涉及语音合成技术领域，方法包括：获得训练文本数据以及真实语音标记数据（步骤110），将训练文本数据以及真实语音标记数据输入至语言处理模型中的预测模块，获得预测模块输出的预测语音标记数据（步骤120），根据真实语音标记数据获得真实风格特征，并根据预测语音标记数据获得预测风格特征（步骤130），根据真实语音标记数据、预测语音标记数据、真实风格特征以及预测风格特征，获得总损失函数（步骤140），根据总损失函数对预测模块进行优化训练（步骤150）。将风格特征损失引入总损失函数，确保语言处理模型能够学习风格信息，使得语言处理模型更准确地复现语音风格，从而提高语言处理模型在处理语音风格信息时的表现，并且语音处理模型是语音合成模型中的关键部分，进而能够提高语音合成大模型在处理语音风格信息时的表现。