

(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 2024/088262 A1

(43) 国际公布日
2024 年 5 月 2 日 (02.05.2024)

- (51) 国际专利分类号:
G10L 15/06 (2013.01) **G10L 15/26** (2006.01)
G10L 15/22 (2006.01)
- (21) 国际申请号: PCT/CN2023/126244
- (22) 国际申请日: 2023 年 10 月 24 日 (24.10.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202211329674.7 2022 年 10 月 27 日 (27.10.2022) CN
- (71) 申请人: 阿里巴巴达摩院 (杭州) 科技有限公司 (ALIBABA DAMO (HANGZHOU) TECHNOLOGY CO., LTD.) [CN/CN]; 中国浙江省杭州市余杭区五常街道文一西路 969 号 3 幢 5 层 516 室, Zhejiang 310023 (CN)。
- (72) 发明人: 周晓欢 (ZHOU, Xiaohuan); 中国北京市朝阳区望京东园四区 6 号楼金辉大厦, Beijing 100102 (CN)。周畅 (ZHOU, Chang); 中国

北京市朝阳区望京东园四区 6 号楼金辉大厦, Beijing 100102 (CN)。

- (74) 代理人: 北京同钧律师事务所 (BEIJING TONGJUN LAW FIRM); 中国北京市海淀区西直门北大街 32 号院 1 号楼 7 层 808 室, Beijing 100082 (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, 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, 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,

(54) **Title:** DATA PROCESSING SYSTEM AND METHOD FOR SPEECH RECOGNITION MODEL, AND SPEECH RECOGNITION METHOD

(54) 发明名称: 语音识别模型的数据处理系统和方法、语音识别方法

(57) **Abstract:** A data processing system and method for a speech recognition model, a speech recognition method, a computing device, and a readable storage medium. The system comprises a cloud-side device 202 and an end-side device 201. The cloud-side device 202 is used for: encoding sample speech data by using an encoder, to obtain speech features of the sample speech data; wherein the encoder is pre-trained on the basis of executing a Chinese pronunciation unit prediction task on pre-training speech data; inputting the speech features into a decoder to obtain predicted Chinese texts, wherein the decoder is pre-trained on the basis of executing a text prediction task on a pre-training Chinese text pronunciation unit; pre-training a model comprising the encoder and the decoder on the basis of the predicted Chinese texts and sample Chinese texts, and when a pre-training stop condition is satisfied, acquiring model parameters of a speech recognition model obtained by pre-training; and sending the model parameters of the speech recognition model to the end-side device 201.

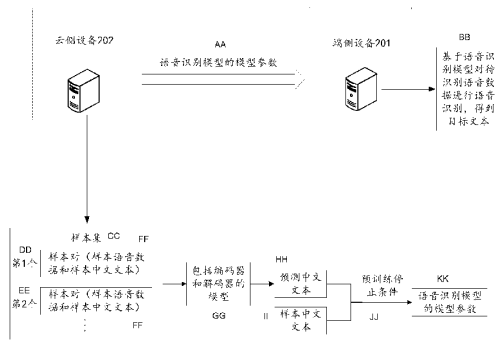


图 2

AA Model parameters of a speech recognition model
BB On the basis of the speech recognition model, perform speech recognition on speech data to be recognized, to obtain a target text
CC Sample set
DD First
EE Second
FF Sample pair (sample speech data and a sample Chinese text)
GG Model comprising an encoder and a decoder
HH Predicted Chinese text
JJ Sample Chinese text
JJ Pre-training stop condition
KK Model parameters of the speech recognition model
LL End-side device
201 End-side device
202 Cloud-side device

[见续页]

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: 一种语音识别模型的数据处理系统及方法、语音识别方法、计算设备及可读存储介质, 该系统包括: 云侧设备202和端侧设备201。云侧设备202用于利用编码器对样本语音数据进行编码, 获得样本语音数据的语音特征, 编码器基于对预训练语音数据执行中文发音单元预测任务进行预训练; 将语音特征输入解码器, 获得预测中文文本, 解码器基于对预训练中文发音单元执行文本预测任务进行预训练; 基于预测中文文本和样本中文文本对包括编码器和解码器的模型进行预训练, 在达到预训练停止条件的情况下, 获取预训练得到的语音识别模型的模型参数; 向端侧设备201发送语音识别模型的模型参数。