

(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
2021 年 11 月 4 日 (04.11.2021)



(10) 国际公布号
WO 2021/218843 A1

(51) 国际专利分类号:
G10L 15/20 (2006.01)

(21) 国际申请号: PCT/CN2021/089556

(22) 国际申请日: 2021 年 4 月 25 日 (25.04.2021)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010366907.5 2020 年 4 月 30 日 (30.04.2020) CN

(71) 申请人: 阿里巴巴集团控股有限公司 (ALIBABA GROUP HOLDING LIMITED) [—/CN]; 开曼群岛大开曼资本大厦一座四层 847 号邮箱, Grand Cayman (KY)。

(72) 发明人: 张仕良 (ZHANG, Shiliang); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。高志付 (GAO, Zhifu); 中国浙江省杭州市余杭区文一西路 969 号 3 号楼 5 楼阿里巴巴集团法务部, Zhejiang 311121 (CN)。

(74) 代理人: 北京三友知识产权代理有限公司 (BEIJING SANYOU INTELLECTUAL PROPERTY AGENCY LTD.); 中国北京市金融街 35 号国际企业大厦 A 座 16 层, Beijing 100033 (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, IT, 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, 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, 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,

(54) Title: STREAMING END-TO-END SPEECH RECOGNITION METHOD AND APPARATUS, AND ELECTRONIC DEVICE

(54) 发明名称: 流式端到端语音识别方法、装置及电子设备

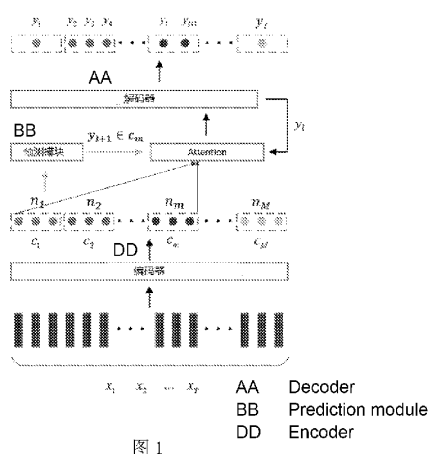


图 1

(57) Abstract: A streaming end-to-end speech recognition method and apparatus, and an electronic device. The method comprises: performing speech acoustic feature extraction and encoding on a received speech stream in a unit of frame (S301); dividing an encoded frame into blocks, and predicting the number of activation points that are comprised in a same block and need to be encoded and output (S302); and determining, according to the prediction result, locations of activation points that need to be decoded and output, so that a decoder can perform decoding at the locations of the activation points and output a recognition result (S303). The method can improve the robustness of a streaming end-to-end speech recognition system in terms of noise, thereby improving the system performance and accuracy.

(57) 摘要: 一种流式端到端语音识别方法、装置及电子设备, 方法包括: 以帧为单位对接收到的语音流进行语音声学特征提取并进行编码 (S301); 对已完成编码的帧进行分块处理, 并对同一分块中包含的需要进行编码输出的激活点数量进行预测 (S302); 根据预测结果确定需要进行解码输出的激活点所在的位置, 以便解码器在激活点所在的位置进行解码并输出识别结果 (S303)。通过本方法能够提升流式端到端语音识别系统对噪声的鲁棒性, 进而提升系统性能以及准确度。

WO 2021/218843 A1

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))。