

(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 2022/151893 A1**

(43) 国际公布日  
2022 年 7 月 21 日 (21.07.2022)

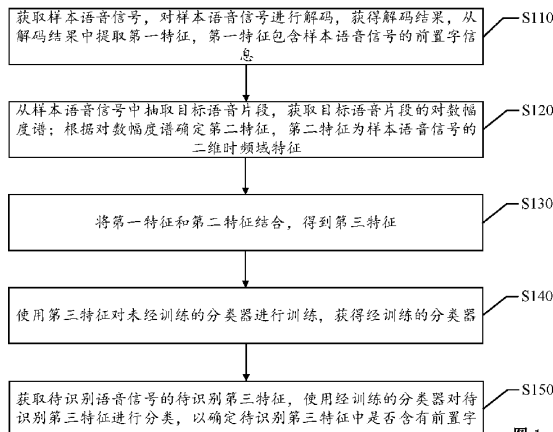
- (51) 国际专利分类号:  
*G10L 15/02* (2006.01) *G10L 15/06* (2013.01)
- (21) 国际申请号: PCT/CN2021/137969
- (22) 国际申请日: 2021 年 12 月 14 日 (14.12.2021)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:  
202110062350.0 2021 年 1 月 18 日 (18.01.2021) CN
- (71) 申请人: 北京沃东天骏信息技术有限公司 (BEIJING WODONG TIANJUN INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。北京京东世纪贸易有限公司 (BEIJING JINGDONG CENTURY TRADING CO., LTD.) [CN/CN]; 中国北京市大兴

区北京经济技术开发区科创十一街 18 号 C 座 2 层 201 室, Beijing 100176 (CN)。

- (72) 发明人: 雪巍(XUE, Wei); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。蔡玉玉(CAI, Yuyu); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。吴俊仪(WU, Junyi); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。彭毅(PENG, Yi); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。范璐(FAN, Lu); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。杨帆(YANG, Fan); 中国北京市北京经济技术开发区科创十一街 18 号院 2 号楼 4 层 A402 室, Beijing 100176 (CN)。丁国宏(DING, Guohong); 中国北京市北京经济技

(54) Title: SPEECH RECOGNITION METHOD AND APPARATUS, STORAGE MEDIUM, AND ELECTRONIC DEVICE

(54) 发明名称: 语音识别方法及装置、存储介质及电子设备



- S110 Obtain a sample speech signal, decode the sample speech signal, obtain a decoding result, and extract a first feature from the decoding result, the first feature comprising prefix information of the sample speech signal
- S120 Extract a target speech segment from the sample speech signal, obtain a log magnitude spectrum of the target speech segment, and determine a second feature according to the log magnitude spectrum, the second feature being a two-dimensional time-frequency domain feature of the sample speech signal
- S130 Combine the first feature and the second feature to obtain a third feature
- S140 Train an untrained classifier by using the third feature so as to obtain a trained classifier
- S150 Obtain a third feature to be recognized of a speech signal to be recognized, and classify, by using the trained classifier, the third feature to be recognized, so as to determine whether the third feature to be recognized comprises a prefix

图 1

(57) Abstract: A speech recognition method, a speech recognition apparatus, a computer readable storage medium, and an electronic device. The method comprises: obtaining a sample speech signal, decoding the sample speech signal, obtaining a decoding result, and extracting a first feature from the decoding result, the first feature comprising prefix information of the sample speech signal (S110); extracting a target speech segment from the sample speech signal, obtaining a log magnitude spectrum of the target speech segment, and determining a second feature according to the log magnitude spectrum, the second feature being a two-dimensional time-frequency domain feature of the sample speech signal (S120); combining the first feature and the second feature to obtain a third feature (S130); training an untrained classifier by using the third feature so as to obtain a trained classifier (S140); and obtaining a third feature to be recognized of a speech signal to be recognized, so as to determine whether the third feature to be recognized comprises a prefix (S150). The present invention provides a method for recognizing a short word comprising a prefix.

[见续页]

WO 2022/151893 A1

术开发区科创十一街18号院2号楼4层A402室, Beijing 100176 (CN)。何晓冬(HE, Xiaodong); 中国北京市北京经济技术开发区科创十一街18号院2号楼4层A402室, Beijing 100176 (CN)。

(74) 代理人: 北京律智知识产权代理有限公司 (BEIJING INTELLEGAL INTELLECTUAL PROPERTY AGENT LTD.); 中国北京市朝阳区慧忠路5号B1605、B1606、B1607, Beijing 100101 (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, 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: 一种语音识别方法、语音识别装置、计算机可读存储介质和电子设备。该方法包括: 获取样本语音信号, 对样本语音信号进行解码, 获得解码结果, 从解码结果中提取第一特征, 第一特征包含样本语音信号的前置字信息(S110); 从样本语音信号中抽取目标语音片段, 获取目标语音片段的对数幅度谱; 根据对数幅度谱确定第二特征, 第二特征为样本语音信号的二维时频域特征(S120); 将第一特征和第二特征结合得到第三特征(S130); 使用第三特征对未经训练的分类器进行训练, 获得经训练的分类器(S140); 获取待识别语音信号的待识别第三特征, 以确定待识别第三特征中是否含有前置字(S150)。该方法提供一种对含有前置字的短字词进行识别的方法。