

(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 2023/082752 A1

(43) 国际公布日
2023 年 5 月 19 日 (19.05.2023)

- (51) 国际专利分类号:
G10L 25/78 (2013.01) *G10L 15/14* (2006.01)
G10L 15/18 (2013.01)
- (21) 国际申请号: PCT/CN2022/113640
- (22) 国际申请日: 2022 年 8 月 19 日 (19.08.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202111337746.8 2021 年 11 月 9 日 (09.11.2021) CN
- (71) 申请人: 京 东 科 技 信 息 有 限 公 司 (JINGDONG TECHNOLOGY 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 6 层 601, Beijing 100176 (CN)。
- (72) 发明人: 王 培 英 (WANG, Peiying); 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 6 层 601, Beijing 100176 (CN)。 杨 久 东

- (YANG, Jiudong); 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 6 层 601, Beijing 100176 (CN)。 陈 蒙 (CHEN, Meng); 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 6 层 601, Beijing 100176 (CN)。
- (74) 代理人: 北 京 清 亦 华 知 识 产 权 代 理 事 务 所 (普 通 合 伙) (TSINGYIHUA INTELLECTUAL PROPERTY LLC); 中国北京市海淀区北洼路 45 号 1 号楼 2 层 201, Beijing 100142 (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, 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,

(54) Title: VOICE DIALOG PROCESSING METHOD AND APPARATUS BASED ON MULTI-MODAL FEATURE, AND ELECTRONIC DEVICE

(54) 发明名称: 基于多模态特征的语音对话处理方法、装置和电子设备

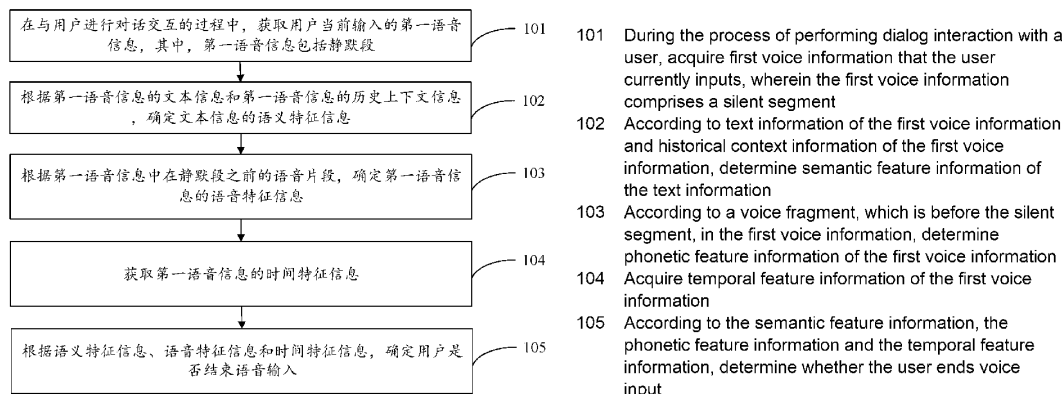


图 1

(57) Abstract: A voice dialog processing method and apparatus (300) based on a multi-modal feature, and an electronic device. The method comprises: during the process of performing dialogue interaction with a user, acquiring first voice information that the user currently inputs, wherein the first voice information comprises a silent segment (101); according to text information of the first voice information and historical context information of the first voice information, determining semantic feature information of the text information (102); according to a voice fragment, which is before the silent segment, in the first voice information, determining phonetic feature information of the first voice information (103); acquiring temporal feature information of the first voice information (104); and according to the semantic feature information, the phonetic feature information and the temporal feature information, determining whether the user ends voice input (105).

[见续页]

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: 一种基于多模态特征的语音对话处理方法、装置(300)和电子设备。方法包括: 在与用户进行对话交互的过程中, 获取用户当前输入的第一语音信息, 其中, 第一语音信息包括静默段(101); 根据第一语音信息的文本信息和第一语音信息的历史上下文信息, 确定文本信息的语义特征信息(102), 根据第一语音信息中在静默段之前的语音片段, 确定第一语音信息的语音特征信息(103), 获取第一语音信息的时间特征信息(104); 根据语义特征信息、语音特征信息和时间特征信息, 确定用户是否结束语音输入(105)。