

(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 4 月 2 日 (02.04.2026)



(10) 国际公布号
WO 2026/066026 A1

(51) 国际专利分类号:
G06F 3/16 (2006.01) **G10L 15/18** (2013.01)
H04R 1/10 (2006.01) **H04W 4/80** (2018.01)

(21) 国际申请号: PCT/CN2025/088880

(22) 国际申请日: 2025 年 4 月 14 日 (14.04.2025)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202411397888.7 2024年9月30日 (30.09.2024) CN

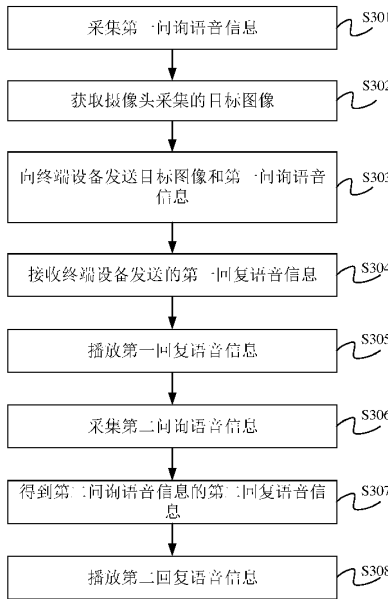
(71) 申请人: 深圳市大十未来科技有限公司 (SHENZHEN DANCING FUTURE TECHNOLOGY LTD.) [CN/CN]; 中国广东省深圳

市南山区沙河街道高发社区深云路二号侨城一号广场主楼2106 518000 (CN)。

(72) 发明人: 蔡玮(CAI, Wei); 中国广东省深圳市南山区沙河街道高发社区深云路二号侨城一号广场主楼2106 518000 (CN)。李新宇(LI, Xinyu); 中国广东省深圳市南山区沙河街道高发社区深云路二号侨城一号广场主楼2106 518000 (CN)。夏清(XIA, Qing); 中国广东省深圳市南山区沙河街道高发社区深云路二号侨城一号广场主楼2106 518000 (CN)。徐萌(XU, Meng); 中国广东省深圳市南山区沙河街道高发社区深云路二号侨城一号广场主楼2106 518000 (CN)。李浩乾(LI, Haoqian); 中国广东省深圳市南山区沙河街道高发社区深云路二号侨城一号广场主楼2106 518000 (CN)。杨炬(YANG, Ju); 中国广东省

(54) Title: VOICE REPLY METHOD, DEVICE, SYSTEM, STORAGE MEDIUM AND PRODUCT

(54) 发明名称: 语音回复方法、设备、系统、存储介质及产品



S301 Collect first query voice information
S302 Acquire a target image collected by a camera
S303 Send to a terminal device the target image and the first query voice information
S304 Receive first reply voice information sent by the terminal device
S305 Playback the first reply voice information
S306 Collect second query voice information
S307 Obtain second reply voice information for the second query voice information
S308 Playback the second reply voice information

图 3

(57) Abstract: The present disclosure relates to the technical field of computers. Provided in the embodiments of the present disclosure are a voice reply method, a device, a system, a storage medium and a product. The method comprises: collecting first query voice information; acquiring a target image collected by a camera; obtaining first reply voice information for the first query voice information, wherein the first reply voice information is information related to the target image, and the target image and the first query voice information are processed by a pre-trained conversational model so as to obtain the first reply voice information; and playing back

[见续页]

深圳市南山区沙河街道高发社区深云路二号
侨城一号广场主楼2106 518000 (CN)。

(74) 代理人: 上海光栅知识产权代理有限公司
(LINKER IP LLC); 中国上海市长宁区延安西
路889号702室 201109 (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, 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))。

the first reply voice information. First, the present application uses earphones to realize intelligent conversation, thus improving the efficiency of intelligent conversation. In addition, the present application can rapidly determine the first reply voice information by using the conversational model. Moreover, replying to the first query voice information by means of acquiring the target image can realize the reply to questions raised by users for real-time environment.

(57) 摘要: 本公开实施例提供一种语音回复方法、设备、系统、存储介质及产品, 涉及计算机技术领域, 该方法包括: 采集第一问询语音信息; 获取摄像头采集的目标图像; 得到第一问询语音信息的第一回复语音信息, 其中, 第一回复语音信息是与目标图像相关的信息, 目标图像和第一问询语音信息被预先训练的对话模型进行处理以得到第一回复语音信息; 播放第一回复语音信息。首先, 本申请采用耳机实现智能对话, 提高了智能对话的效率。此外, 本申请采用对话模型能够快速的确定的第一回复语音信息。再者, 通过获取目标图像以回复第一问询语音信息, 可以实现对用户针对实时的环境提出的问题的回复。