

(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 2024/017110 A1

(43) 国际公布日
2024 年 1 月 25 日 (25.01.2024)

(51) 国际专利分类号:

G10L 21/0208 (2013.01) **G10L 21/0232** (2013.01)
G10L 21/0216 (2013.01) **G10L 25/30** (2013.01)
G10L 21/0224 (2013.01)

(21) 国际申请号: PCT/CN2023/106951

(22) 国际申请日: 2023 年 7 月 12 日 (12.07.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202210864010.4 2022 年 7 月 21 日 (21.07.2022) CN

(71) 申请人: 广州市百果园信息技术有限公司 (GUANGZHOU BAIGUOYUAN INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国广东省广州市番禺区南村镇万博二路 79 号 201-203 房、401-410 房、801-811 房、1001-1011 房、1101-1111 房, Guangdong 511402 (CN)。

(72) 发明人: 魏善义 (WEI, Shanyi); 中国广东省广州市番禺区南村镇万博二路 79 号 201-203 房、

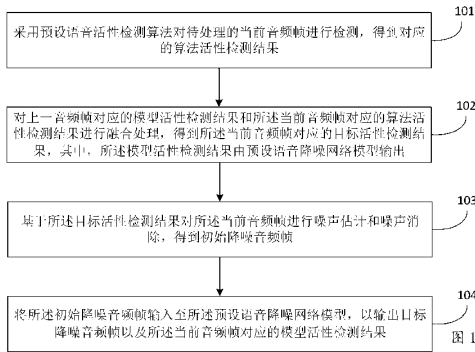
401-410 房、801-811 房、1001-1011 房、1101-1111 房, Guangdong 511402 (CN)。刘梁 (LIU, Liang); 中国广东省广州市番禺区南村镇万博二路 79 号 201-203 房、401-410 房、801-811 房、1001-1011 房、1101-1111 房, Guangdong 511402 (CN)。

(74) 代理人: 北京品源专利代理有限公司 (BEYOND ATTORNEYS AT LAW); 中国北京市海淀区莲花池东路 39 号西金大厦 6 层, Beijing 100036 (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。

(54) Title: VOICE NOISE REDUCTION METHOD, MODEL TRAINING METHOD, APPARATUS, DEVICE, MEDIUM, AND PRODUCT

(54) 发明名称: 语音降噪方法、模型训练方法、装置、设备、介质及产品



(57) Abstract: A voice noise reduction method, a model training method, an apparatus, a device, a medium, and a product. The voice noise reduction method comprises: using a preset voice activity detection algorithm to detect a current audio frame to be processed, and obtaining a corresponding algorithm activity detection result [101]; merging a model activity detection result corresponding to the previous audio frame with an algorithm activity detection result corresponding to the current audio frame, obtaining a target activity detection result corresponding to the current audio frame, wherein the model activity detection result is outputted by a preset voice noise reduction network model [102]; based on the target activity detection result, performing noise estimation and noise elimination on the current audio frame, obtaining an initial noise reduction audio frame [103]; and inputting the initial noise reduction audio frame into the preset voice noise reduction network model so as to output a target noise reduction audio frame and also a model activity detection result corresponding to the current audio frame [104]. The use of the aforementioned solution can enhance a voice noise reduction effect, and the stability and robustness of the voice noise reduction solution are improved.

- 101 Use a preset voice activity detection algorithm to detect a current audio frame to be processed, and obtain a corresponding algorithm activity detection result
- 102 Merge a model activity detection result corresponding to the previous audio frame with an algorithm activity detection result corresponding to the current audio frame, and obtain a target activity detection result corresponding to the current audio frame, wherein the model activity detection result is outputted by a preset voice noise reduction network model
- 103 Based on the target activity detection result, perform noise estimation and noise elimination on the current audio frame, and obtain an initial noise reduction audio frame
- 104 Input the initial noise reduction audio frame into the preset voice noise reduction network model so as to output a target noise reduction audio frame and also a model activity detection result corresponding to the current audio frame

图 1

WO 2024/017110 A1

[见续页]

(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))。

(57) 摘要: 一种语音降噪方法、模型训练方法、装置、设备、介质及产品。其中, 该语音降噪方法包括: 采用预设语音活性检测算法对待处理的当前音频帧进行检测, 得到对应的算法活性检测结果[101]; 对上一音频帧对应的模型活性检测结果和当前音频帧对应的算法活性检测结果进行融合处理, 得到当前音频帧对应的目标活性检测结果, 模型活性检测结果由预设语音降噪网络模型输出[102]; 基于目标活性检测结果对当前音频帧进行噪声估计和噪声消除, 得到初始降噪音频帧[103]; 将初始降噪音频帧输入至预设语音降噪网络模型, 以输出目标降噪音频帧以及当前音频帧对应的模型活性检测结果[104]。通过采用上述方案, 可以提升语音降噪效果, 并提高语音降噪方案的稳定性和鲁棒性。