

(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 2024/103302 A1

(43) 国际公布日
2024 年 5 月 23 日 (23.05.2024)

WIPO | PCT

- (51) 国际专利分类号:
G10H 1/00 (2006.01) **G10L 25/30** (2013.01)
G10L 21/0272 (2013.01)
- (21) 国际申请号: PCT/CN2022/132325
- (22) 国际申请日: 2022 年 11 月 16 日 (16.11.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (71) 申请人: 广州酷狗计算机科技有限公司 (GUANGZHOU KUGOU COMPUTER TECHNOLOGY CO., LTD.) [CN/CN]; 中国广东省广州市天河区黄埔大道中 315 号自编 1-17, Guangdong 510660 (CN)。
- (72) 发明人: 罗程方 (LUO, Chengfang); 中国广东省广州市天河区黄埔大道中 315 号自编 1-17, Guangdong

510660 (CN)。万景轩 (WAN, Jingxuan); 中国广东省广州市天河区黄埔大道中 315 号自编 1-17, Guangdong 510660 (CN)。陈传艺 (CHEN, Chuanyi); 中国广东省广州市天河区黄埔大道中 315 号自编 1-17, Guangdong 510660 (CN)。

(74) 代理人: 北京三高永信知识产权代理有限公司 (BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市海淀区上地信息产业基地三街 1 号楼四层 C 段 457, Beijing 100085 (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,

(54) Title: HUMAN VOICE NOTE RECOGNITION MODEL TRAINING METHOD, HUMAN VOICE NOTE RECOGNITION METHOD, AND DEVICE

(54) 发明名称: 人声音符识别模型的训练方法、人声音符识别方法及设备

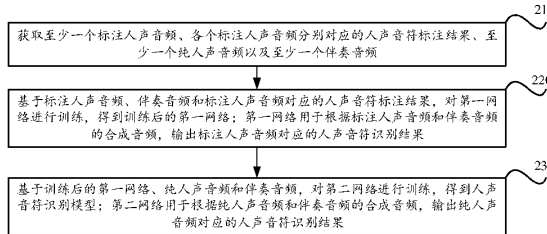


图 2

- 210 Acquire at least one labeled human voice audio, a human voice note labeled result respectively corresponding to each labeled human voice audio, at least one pure human voice audio, and at least one accompaniment audio
- 220 Train a first network on the basis of the labeled human voice audio, the accompaniment audio, and the human voice note labeled result corresponding to the labeled human voice audio to obtain a trained first network, the first network being used for outputting, according to a synthesized audio of the labeled human voice audio and the accompaniment audio, a human voice note recognition result corresponding to the labeled human voice audio
- 230 Train a second network on the basis of the trained first network, the pure human voice audio, and the accompaniment audio to obtain a human voice note recognition model, the second network being used for outputting, according to a synthesized audio of the pure human voice audio and the accompaniment audio, a human voice note recognition result corresponding to the pure human voice audio

(57) Abstract: A human voice note recognition model training method, a human voice note recognition method, and a device, relating to the technical field of artificial intelligence. The method comprises: acquiring at least one labeled human voice audio, a human voice note labeled result respectively corresponding to each labeled human voice audio, at least one pure human voice audio, and at least one accompaniment audio; training a first network on the basis of the labeled human voice audio, the accompaniment audio, and the human voice note labeled result corresponding to the labeled human voice audio to obtain a trained first network; and training a second

[见续页]

WO 2024/103302 A1

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,
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, 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))。

network on the basis of the trained first network, the pure human voice audio, and the accompaniment audio to obtain a human voice note recognition model. According to the obtained human voice note recognition model, a human voice accompaniment separation algorithm does not need to be called, thereby reducing the calculation complexity of human voice note recognition.

(57) 摘要: 一种人声音符识别模型的训练方法、人声音符识别方法及设备, 涉及人工智能技术领域。
上述方法包括: 获取至少一个标注人声音频、各个标注人声音频分别对应的人声音符标注结果、至少一个纯人声音频以及至少一个伴奏音频; 基于标注人声音频、伴奏音频和标注人声音频对应的人声音符标注结果, 对第一网络进行训练, 得到训练后的第一网络; 基于训练后的第一网络、纯人声音频和伴奏音频, 对第二网络进行训练, 得到人声音符识别模型。得到的人声音符识别模型, 无需调用人声伴奏分离算法, 降低了人声音符识别的计算复杂度。