

(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
2022 年 9 月 9 日 (09.09.2022)



(10) 国际公布号
WO 2022/183968 A1

(51) 国际专利分类号:
G10L 17/18 (2013.01)

中国浙江省杭州市滨江区长河街道网商路 699 号 4 号楼 5 楼 508 室, Zhejiang 310052 (CN)。

(21) 国际申请号: PCT/CN2022/077742

(74) 代理人: 北京博浩百睿知识产权代理有限公司 (KSK IP AGENCY); 中国北京市海淀区知春路甲 48 号盈都大厦 C 座 1 单元 10E, Beijing 100098 (CN)。

(22) 国际申请日: 2022 年 2 月 24 日 (24.02.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110235834.0 2021 年 3 月 3 日 (03.03.2021) 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, 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, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(71) 申请人: 阿里巴巴 (中国) 有限公司 (ALIBABA (CHINA) CO., LTD.) [CN/CN]; 中国浙江省杭州市滨江区长河街道网商路 699 号 4 号楼 5 楼 508 室, Zhejiang 310052 (CN)。

(72) 发明人: 郑斯奇 (ZHENG, Siqi); 中国浙江省杭州市滨江区长河街道网商路 699 号 4 号楼 5 楼 508 室, Zhejiang 310052 (CN)。 索宏彬 (SUO, Hongbin);

(54) Title: AUDIO SIGNAL PROCESSING METHOD, DEVICES, SYSTEM, AND STORAGE MEDIUM

(54) 发明名称: 音频信号处理方法、设备、系统及存储介质

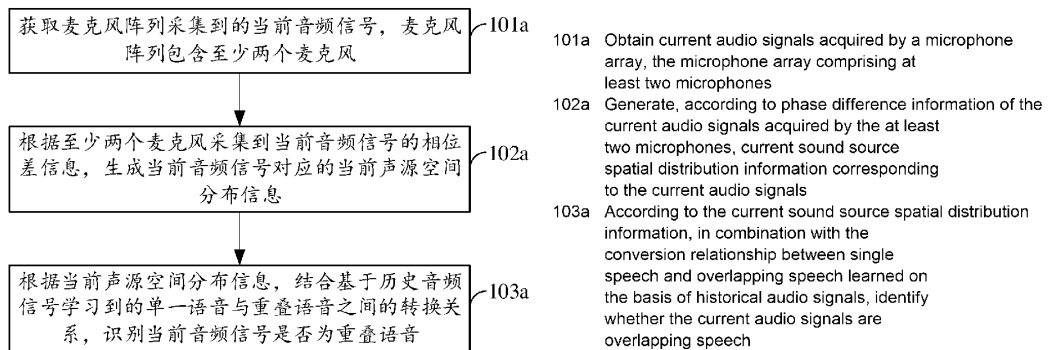


图 1a

(57) Abstract: An audio signal processing method, a system, a terminal device, a conference device, a teaching device, an intelligent vehicle-mounted device, a server device, and a computer-readable storage medium. The method comprises: obtaining current audio signals acquired by a microphone array, the microphone array comprising at least two microphones (101a); generating, according to phase difference information of the current audio signals acquired by the at least two microphones, current sound source spatial distribution information corresponding to the current audio signals (102a); and according to the current sound source spatial distribution information, in combination with the conversion relationship between single speech and overlapping speech learned on the basis of historical audio signals, identifying whether the current audio signals are overlapping speech (103a). Compared with single-channel audio, the audio signals acquired by the microphone array are used, and the sound source spatial distribution information is included, thus, it can be accurately identified whether the current audio signals are overlapping speech, thereby satisfying the detection requirement for a product level.

[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, 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: 一种音频信号处理方法、系统、终端设备、会议设备、教学设备、智能车载设备、服务端设备及计算机可读存储介质, 其中该方法包括: 获取麦克风阵列采集到的当前音频信号, 麦克风阵列包含至少两个麦克风 (101a); 根据至少两个麦克风采集到当前音频信号的相位差信息, 生成当前音频信号对应的当前声源空间分布信息(102a); 根据当前声源空间分布信息, 结合基于历史音频信号学习到的单一语音与重叠语音之间的转换关系, 识别当前音频信号是否为重叠语音(103a)。相比于单通道音频, 利用麦克风阵列采集的音频信号, 包含了声源空间分布信息, 可以准确地识别当前音频信号是否为重叠语音, 满足产品级别的检测需求。