

(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 2019/227672 A1

(43) 国际公布日
2019 年 12 月 5 日 (05.12.2019)

(51) 国际专利分类号:
G10L 17/00 (2013.01)

(21) 国际申请号: PCT/CN2018/100174

(22) 国际申请日: 2018 年 8 月 13 日 (13.08.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810519521.6 2018 年 5 月 28 日 (28.05.2018) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(72) 发明人: 赵峰(ZHAO, Feng); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
王健宗(WANG, Jianzong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。
肖京(XIAO, Jing); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市赛恩倍吉知识产权代理有限公司(SHENZHEN SCIENBIZIP INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市龙华新区龙观东路 83 号荣群大厦 9 楼, Guangdong 518109 (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,

(54) Title: VOICE SEPARATION MODEL TRAINING METHOD, TWO-SPEAKER SEPARATION METHOD AND ASSOCIATED APPARATUS

(54) 发明名称: 说话人分离模型训练方法、两说话人分离方法及相关设备

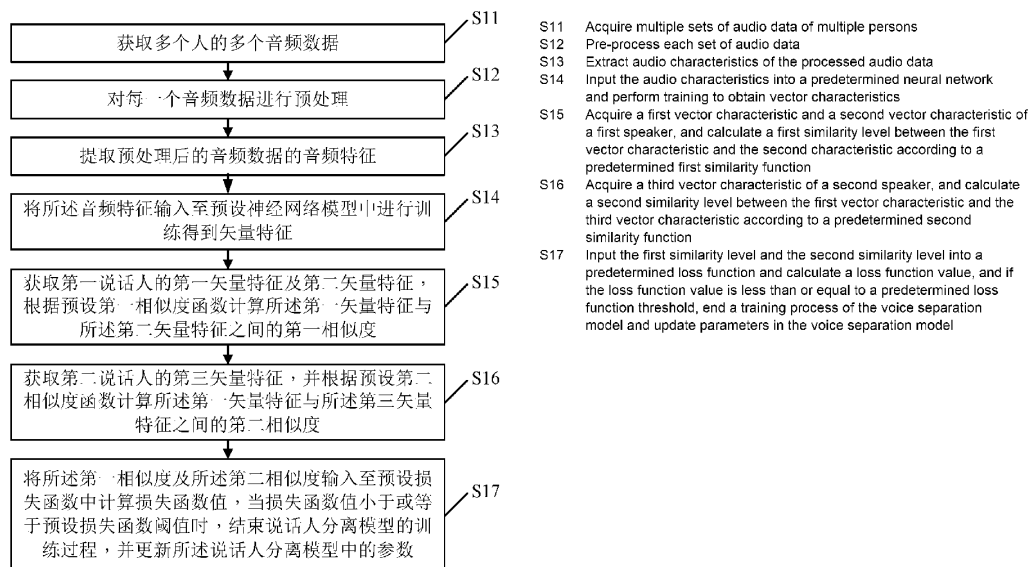


图 1

(57) Abstract: A voice separation model training method comprises: acquiring and pre-processing multiple sets of audio data; extracting audio characteristics of pre-processed audio data; inputting the audio characteristics into a predetermined neural network model to obtain vector characteristics; calculating a first similarity level between a first vector characteristic and a second vector characteristic of a first speaker; calculating a second similarity level between the first vector characteristic level of the first speaker and a third vector characteristic of a second speaker; calculating a loss function value of the first similarity level and the second similarity level, and if

[见续页]

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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。

the loss function value is less than or equal to a predetermined loss function threshold, ending a training process of the voice separation model and updating parameters in the model. The present application further provides a two-speaker separation method, a terminal and a storage medium. The present application is capable of significantly enhancing characteristic extraction capabilities of a model for input voice data, improving separation accuracy for two speakers, and more particularly achieving better separation results for a separation task of a lengthy dialogue.

(57) 摘要: 一种说话人分离模型训练方法, 包括: 获取多个音频数据并进行预处理; 提取预处理后的音频数据的音频特征; 将音频特征输入至预设神经网络模型中进行训练得到矢量特征; 计算第一说话人的第一矢量特征与第二矢量特征之间的第一相似度; 计算第一说话人的第一矢量特征与第二说话人的第三矢量特征之间的第二相似度; 计算第一相似度及第二相似度的损失函数值, 当损失函数值小于或等于预设损失函数阈值时, 结束说话人分离模型的训练过程, 并更新模型中的参数。本申请还提供一种两说话人分离方法、终端及存储介质。能够显著的增强模型对输入的语音数据的特征提取能力, 提高两说话人分离的准确率, 尤其是在长时对话的分离任务中能取得更佳的分隔效果。