

(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 2022/253061 A1

(43) 国际公布日
2022 年 12 月 8 日 (08.12.2022)

(51) 国际专利分类号:
G10L 13/033 (2013.01) *G10L 25/48* (2013.01)
G10L 13/10 (2013.01) *G10L 25/30* (2013.01)

(21) 国际申请号: PCT/CN2022/094838

(22) 国际申请日: 2022 年 5 月 25 日 (25.05.2022)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202110621213.6 2021 年 6 月 3 日 (03.06.2021) CN

(71) 申请人: 华为技术有限公司 (HUAWEI TECHNOLOGIES CO., LTD.) [CN/CN]; 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(72) 发明人: 邓利群 (DENG, Liqun); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 谭达新 (TAN, Daxin); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。 郑念祖 (ZHENG, Nianzu); 中国广东省深圳市龙岗区坂田华为总部办公楼, Guangdong 518129 (CN)。

(74) 代理人: 深圳市深佳知识产权代理事务所 (普通合伙) (SHENPAT INTELLECTUAL PROPERTY AGENCY); 中国广东省深圳市罗湖区南湖街道春风路庐山大厦 B 座 18C2、18D、18E、18E2, Guangdong 518001 (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,

(54) Title: VOICE PROCESSING METHOD AND RELATED DEVICE

(54) 发明名称: 一种语音处理方法及相关设备

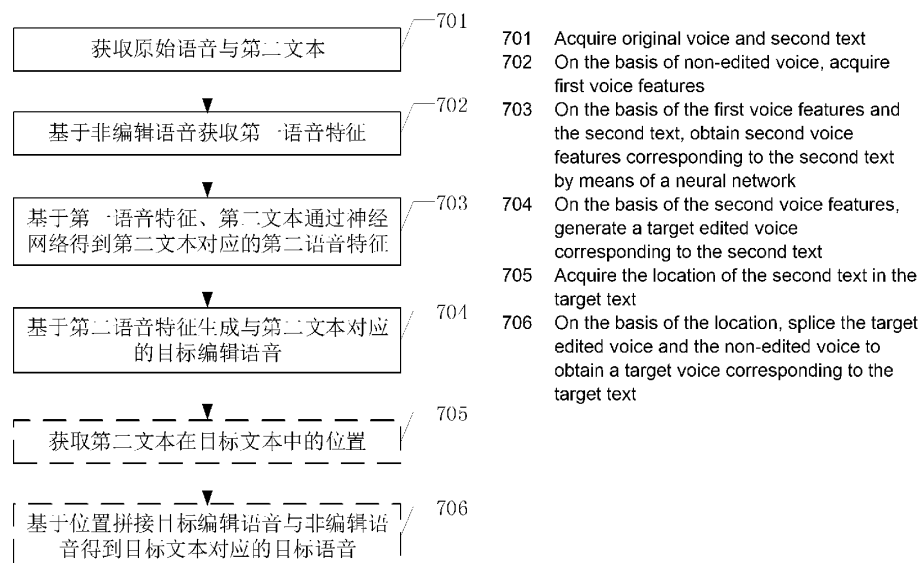


图 7

(57) Abstract: A voice processing method and a related device, applicable to scenarios such as a user recording a short video and a teacher recording teaching voice, the method comprising: acquiring original voice and second text (701), original text corresponding to the original voice and target text to which the second text belongs both comprising first text, and voice corresponding to the first text in the original voice being non-edited voice; on the basis of the non-edited voice, acquiring first voice features (702); on the basis of the first voice features and the second text, obtaining second voice features corresponding to the second text by means of a neural network (703); and, on the basis of the second voice features, generating a target edited voice corresponding to the second text (704).

[见续页]

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, 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。

- (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 present method enables the sensation of listening to the target edited voice corresponding to a modified part to be similar to the sensation of listening to the non-edited voice corresponding to the correct text, enhancing the user experience.

(57) 摘要: 一种语音处理方法及相关设备, 可以应用于用户录制短视频、老师录制授课语音等场景, 该方法包括: 获取原始语音与第二文本(701), 原始语音对应的原始文本与第二文本所属的目标文本都包括第一文本, 第一文本在原始语音中对应的语音为非编辑语音; 基于非编辑语音获取第一语音特征(702); 基于第一语音特征与第二文本通过神经网络得到第二文本对应的第二语音特征(703); 基于第二语音特征生成第二文本对应的目标编辑语音(704)。该方法实现了修改部分对应的目标编辑语音的听感与正确文本对应的非编辑语音的听感类似, 提升用户体验。