

(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 2024/109375 A1

(43) 国际公布日
2024 年 5 月 30 日 (30.05.2024)

(51) 国际专利分类号:

G10L 13/08 (2013.01) *G10L 15/06* (2013.01)
G10L 15/26 (2006.01)

(21) 国际申请号: PCT/CN2023/124162

(22) 国际申请日: 2023 年 10 月 12 日 (12.10.2023)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:

202211455842.7 2022 年 11 月 21 日 (21.11.2022) CN

(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(72) 发明人: 杨培基 (YANG, Peiji); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (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, 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。

(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

(54) Title: METHOD AND APPARATUS FOR TRAINING SPEECH CONVERSION MODEL, DEVICE, AND MEDIUM

(54) 发明名称: 语音转换模型的训练方法、装置、设备及介质

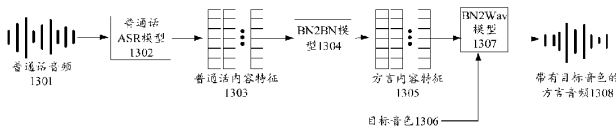


图 13

1301 Mandarin audio
1302 Mandarin ASR model
1303 Mandarin content feature
1304 BN2BN model
1305 Dialect content feature
1306 Target timbre
1307 BN2Wav model
1308 Dialect audio having the target timbre

(57) Abstract: A method and apparatus for training a speech conversion model, a device, and a medium. The method comprises: training a first ASR model on the basis of first sample audio, and training a second ASR model on the basis of second sample audio (201); training a first conversion model on the basis of a first sample text and a first sample content feature corresponding to the first sample audio (202); constructing parallel sample data on the basis of the first conversion model, and a second sample text and a second sample content feature corresponding to the second sample audio (203); training a second conversion model on the basis of the parallel sample data, the second conversion model being used to perform content feature conversion between a first accent and a second accent (204); training a third conversion model on the basis of sample content features of different sample audios (205); generating a speech conversion model on the basis of the trained first ASR model, second conversion model and third conversion model (206).

(57) 摘要: 一种语音转换模型的训练方法、装置、设备及介质。包括: 基于第一样本音频训练第一 ASR 模型, 以及基于第二样本音频训练第二 ASR 模型(201); 基于第一样本音频对应的第一样本文本以及第一样本内容特征, 训练第一转换模型(202); 基于第一转换模型、第二样本音频对应的第二样本文本以及第二样本内容特征, 构建平行样本数据(203); 基于平行样本数据训练第二转换模型, 第二转换模型用于对第一口音和第二口音间进行内容特征转换(204); 基于不同样本音频的样本内容特征训练第三转换模型(205); 基于训练得到的第一 ASR 模型、第二转换模型和第三转换模型生成语音转换模型(206)。

WO 2024/109375 A1

[见续页]

(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))。