

(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 2022/083083 A1

(43) 国际公布日
2022 年 4 月 28 日 (28.04.2022)

(51) 国际专利分类号:

G10L 15/06 (2013.01)

G10L 25/24 (2013.01)

G10L 15/02 (2006.01)

G10L 25/30 (2013.01)

G10L 19/16 (2013.01)

南京市雨花台区凤信路20号万博科技园C号楼4层, Jiangsu 210012 (CN)。

(21) 国际申请号:

PCT/CN2021/088507

(22) 国际申请日:

2021 年 4 月 20 日 (20.04.2021)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202011129857.5 2020年10月21日 (21.10.2020) CN

(72) 发明人: 司马华鹏(SIMA, Huapeng); 中国江苏省南京市雨花台区凤信路20号万博科技园C号楼4层, Jiangsu 210012 (CN)。 毛志强(MAO, Zhiqiang); 中国江苏省南京市雨花台区凤信路20号万博科技园C号楼4层, Jiangsu 210012 (CN)。 龚雪飞(GONG, Xuefei); 中国江苏省南京市雨花台区凤信路20号万博科技园C号楼4层, Jiangsu 210012 (CN)。

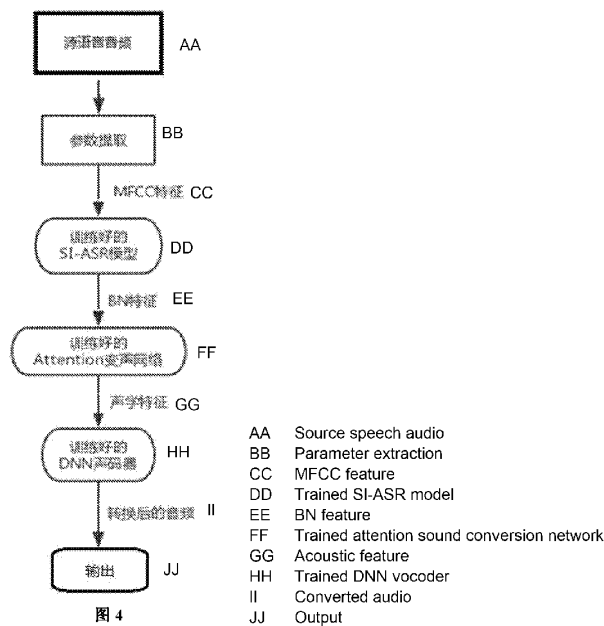
(71) 申请人: 南京硅基智能科技有限公司 (NANJING SILICON INTELLIGENCE TECHNOLOGY CO., LTD.) [CN/CN]; 中国江苏省

(74) 代理人: 北京弘权知识产权代理有限公司 (CHINABLE IP); 中国北京市朝阳区安定路35号六层35-10-2内620室, Beijing 100029 (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: SOUND CONVERSION SYSTEM AND TRAINING METHOD FOR SAME

(54) 发明名称: 一种声音变换系统以及声音变换系统的训练方法



(57) Abstract: A speech conversion solution for non-parallel corpus training gets rid of the dependence on a parallel text, and solves the technical problem that speech conversion is difficult to achieve under the condition that resources and devices are limited. The present invention comprises a sound conversion system and a training method for same. A trained speaker-independent speech recognition model can be used for any source speaker, that is, a speaker is independent; and a bottleneck feature of an audio is more abstract than a phonetic posteriorgram feature, not only can reflect speech content and decouple from the timbre of the speaker, but also is not so

[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, IT, 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))。

closely bound to a phoneme category, and does not have a clear one-to-one correspondence, thereby relieving, to a certain extent, the problem of inaccurate pronunciation caused by an ASR recognition error. The pronunciation accuracy of an audio obtained by performing sound conversion using the bottleneck feature is significantly higher than the pronunciation accuracy of an audio obtained by means of a phonetic posteriorgram method, and there is no significant difference in timbre; and using a transfer learning mode can greatly reduce the dependence on training corpus.

(57) 摘要: 一种非平行语料训练的语音转换方案, 摆脱对平行文本的依赖, 解决在资源、设备有限的条件下难以实现语音转换的技术问题, 包括一种声音变换系统及声音变换系统的训练方法, 训练好的说话人独立的语音识别模型可用于任意的源说话人, 即说话人独立; 音频的瓶颈特征相比语音后验概率特征更加抽象, 既能反映说话内容和说话人音色解耦, 同时又和音素类别的绑定没有那么紧密, 不是明确的一一对应关系, 一定程度上缓解了ASR识别错误导致发音不准的问题。使用瓶颈特征作声音变换得到的音频, 发音的准确率明显高于语音后验概率方法, 并且音色没有显著区别; 利用迁移学习方式, 可以大幅度缩小训练语料的依赖。