

(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 2023/207472 A1

(43) 国际公布日
2023 年 11 月 2 日 (02.11.2023)

- (51) 国际专利分类号:
G10L 13/02 (2013.01) *G10L 13/033* (2013.01)
- (21) 国际申请号: PCT/CN2023/084236
- (22) 国际申请日: 2023 年 3 月 28 日 (28.03.2023)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202210460941.8 2022 年 4 月 28 日 (28.04.2022) CN
- (71) 申请人: 腾讯音乐娱乐科技(深圳)有限公司 (TENCENT MUSIC ENTERTAINMENT TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市前海深港合作区前湾一

- 路 1 号 A 栋 201 室 (入驻深圳市前海商务秘书有限公司), Guangdong 518052 (CN)。
- (72) 发明人: 徐东 (XU, Dong); 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室 (入驻深圳市前海商务秘书有限公司), Guangdong 518052 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI,

(54) Title: AUDIO SYNTHESIS METHOD, ELECTRONIC DEVICE AND READABLE STORAGE MEDIUM

(54) 发明名称: 一种音频合成方法、电子设备及可读存储介质

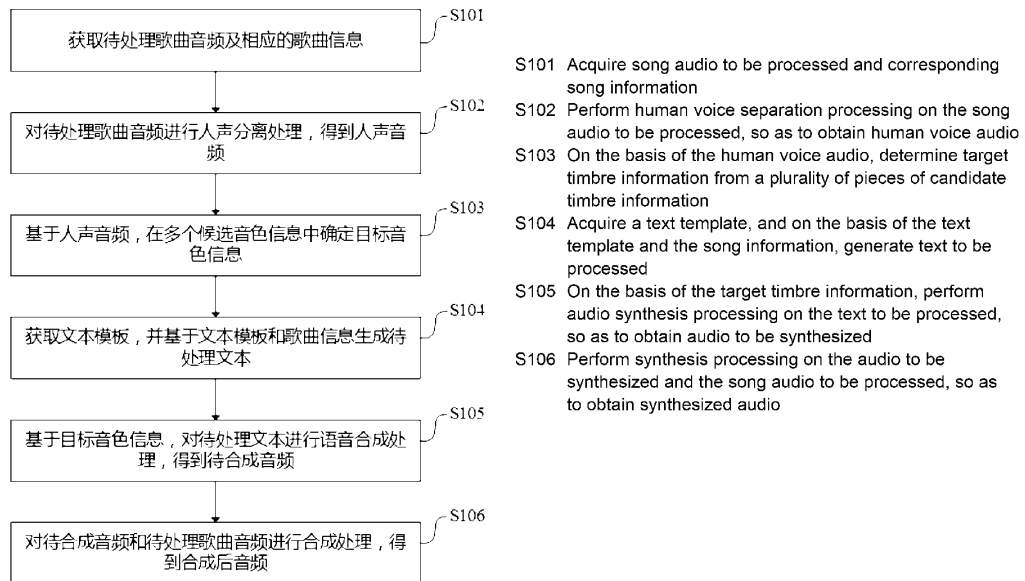


图 3

(57) Abstract: An audio synthesis method, a device and a computer-readable storage medium. The method comprises: acquiring song audio to be processed and corresponding song information (S101); performing human voice separation processing on the song audio to be processed, so as to obtain human voice audio (S102); on the basis of the human voice audio, determining target timbre information from a plurality of pieces of candidate timbre information (S103); acquiring a text template, and on the basis of the text template and the song information, generating text to be processed (S104); on the basis of the target timbre information, performing audio synthesis

[见续页]

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 (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))。

processing on the text to be processed, so as to obtain audio to be synthesized (S105); and performing synthesis processing on the audio to be synthesized and the song audio to be processed, so as to obtain synthesized audio (S106). The quality of synthesized audio which is obtained by means of the method is not limited to the artificial broadcasting level, such that the quality of synthesized audio is higher, and the playing effect is better.

(57) 摘要: 一种音频合成方法、设备及计算机可读存储介质, 该方法包括: 获取待处理歌曲音频及相应的歌曲信息 (S101); 对待处理歌曲音频进行人声分离处理, 得到人声音频 (S102); 基于人声音频, 在多个候选音色信息中确定目标音色信息 (S103); 获取文本模板, 并基于文本模板和歌曲信息生成待处理文本 (S104); 基于目标音色信息, 对待处理文本进行语音合成处理, 得到待合成音频 (S105); 对待合成音频和待处理歌曲音频进行合成处理, 得到合成后音频 (S106); 该方法得到的合成后音频的质量不受限于人工播音水平, 使得合成后音频的质量较高, 播放效果较好。