

(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 2025/175863 A1

(43) 国际公布日
2025 年 8 月 28 日 (28.08.2025)

- (51) 国际专利分类号:
H04N 21/439 (2011.01) **G10L 13/033** (2013.01)
H04N 21/472 (2011.01)
- (21) 国际申请号: PCT/CN2024/135302
- (22) 国际申请日: 2024 年 11 月 28 日 (28.11.2024)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202410185739.8 2024年2月19日 (19.02.2024) CN
- (71) 申请人: 北京字跳网络技术有限公司
(**BEIJING ZITIAO NETWORK TECHNOLOGY CO., LTD.**) [CN/CN]; 中国北京市海淀区紫金数码园4号楼2层0207 100190 (CN)。
- (72) 发明人: 楚书廷(**CHU, Shuting**); 中国北京市朝阳区七圣中街12号院融中心B1小邮局 100028 (CN)。

李忠豪(**LI, Zhonghao**); 中国北京市朝阳区七圣中街12号院融中心B1小邮局 100028 (CN)。

(74) 代理人: 北京世辉律师事务所(**SHIHUI PARTNERS**); 中国北京市朝阳区建国门外大街2号北京银泰中心C座42层 100022 (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。

(54) **Title:** AUDIO PROCESSING METHOD AND APPARATUS, AND DEVICE AND STORAGE MEDIUM

(54) 发明名称: 音频处理的方法、装置、设备和存储介质

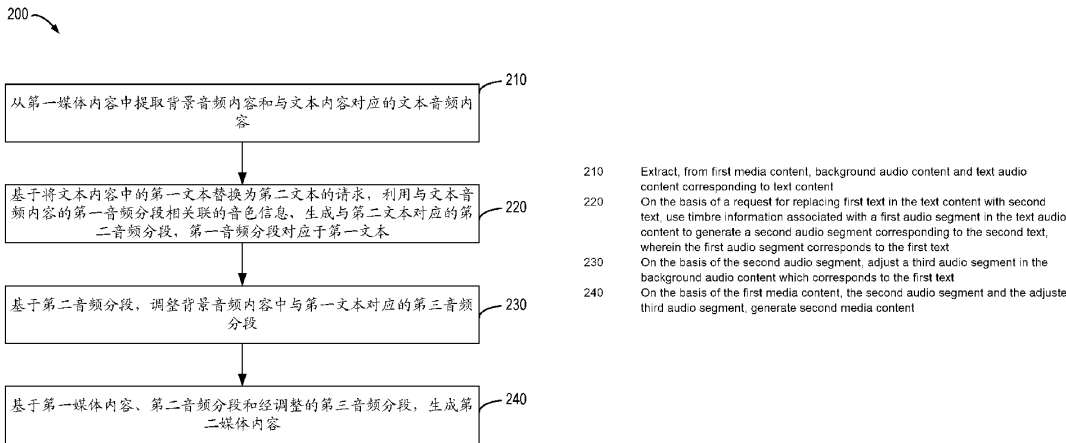


图 2

(57) **Abstract:** An audio processing method and apparatus, and a device and a storage medium. The method comprises: extracting, from first media content, background audio content and text audio content corresponding to text content (210); on the basis of a request for replacing first text in the text content with second text, using timbre information associated with a first audio segment in the text audio content to generate a second audio segment corresponding to the second text, wherein the first audio segment corresponds to the first text (220); on the basis of the second audio segment, adjusting a third audio segment in the background audio content which corresponds to the first text (230); and on the basis of the first media content, the second audio segment and the adjusted third audio segment, generating second media content (240). In this way, the method can support editing media content by means of modifying text corresponding to the media content, and can make the modified audio content more real, thereby improving the quality of the edited media content.

[见续页]

(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))。

(57) 摘要: 一种音频处理的方法、装置、设备和存储介质, 该方法包括: 从第一媒体内容中提取背景音频内容和与文本内容对应的文本音频内容 (210); 基于将文本内容中的第一文本替换为第二文本的请求, 利用与文本音频内容的第一音频分段相关联的音色信息, 生成与第二文本对应的第二音频分段, 第一音频分段对应于第一文本 (220); 基于第二音频分段, 调整背景音频内容中与第一文本对应的第三音频分段 (230); 以及基于第一媒体内容、第二音频分段和经调整的第三音频分段, 生成第二媒体内容 (240)。通过此方式, 该方法能够支持通过修改媒体内容对应的文本来实现对媒体内容的编辑, 并能够使得修改后的音频内容更为真实, 从而提高编辑后的媒体内容的质量。