

(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

(43) 国际公布日
2023 年 1 月 19 日 (19.01.2023)



(10) 国际公布号
WO 2023/287360 A2

- (51) 国际专利分类号:
无分类
- (21) 国际申请号: PCT/SG2022/050494
- (22) 国际申请日: 2022 年 7 月 14 日 (14.07.2022)
- (25) 申请语言: 中文
- (26) 公布语言: 中文
- (30) 优先权:
202110802038.0 2021 年 7 月 15 日 (15.07.2021) CN
- (71) 申请人: 脸萌有限公司 (LEMON INC.) [GB/SG]; 开曼群岛大开曼西湾路 802 号木槿道大楼邮政信箱 31119, Grand Cayman KY1 - 1205 (KY)。
- (72) 发明人: 郑鑫 (ZHENG, Xin); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。朱聪慧 (ZHU, Conghui); 美国加利福尼亚州洛杉矶西杰弗逊大道 12655 号六层第 137 室, California 90066 (US)。夏瑞 (XIA, Rui); 美国加利福尼亚州洛杉矶西杰弗逊大道 12655 号六层第 137 室, California 90066 (US)。尚

楚翔 (SHANG, Chuxiang); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。钟德建 (ZHONG, Dejian); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。蒋泳森 (JIANG, Yongsen); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。屠明 (TU, Ming); 美国加利福尼亚州洛杉矶西杰弗逊大道 12655 号六层第 137 室, California 90066 (US)。邓乐来 (DENG, Lelai); 中国北京市海淀区知春路 63 号中国卫星通信大厦今日头条小邮局, Beijing 100086 (CN)。

(74) 代理人: 傅子健 (POH, Chee Kian, Daniel); 新加坡新加坡麦仕奇新加坡, 丹戎巴葛邮局, P O Box 636, Singapore 910816 (SG)。

(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,

(54) Title: MULTIMEDIA PROCESSING METHOD AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 多媒体处理方法、装置、电子设备及存储介质

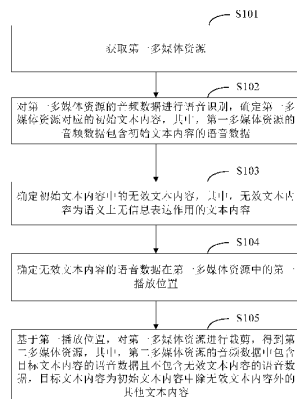


图 3

(57) Abstract: Provided in the embodiments of the present disclosure are a multimedia processing method and apparatus, an electronic device, and a storage medium. The method comprises: acquiring a first multimedia resource; performing voice recognition on audio data in the first multimedia resource to determine an initial text content corresponding to the first multimedia resource, wherein the audio data of the first multimedia resource includes voice data of the initial text content; determining invalid text content in the initial text content, wherein the invalid text content is text content that semantically does not express information; determining a first playing position of the voice data of the invalid text content in the first multimedia resource; and on the basis of the first playing position, cutting the first multimedia resource, so as to obtain a second multimedia resource, wherein audio data of the second multimedia resource includes the voice data of the target text content and does not include the voice data of the invalid text content. By means of the embodiments of the present disclosure, invalid content in the multimedia resource can be automatically cut, and the cutting efficiency and the cutting effect are improved.

- S101 Acquire a first multimedia resource
- S102 Perform voice recognition on audio data in the first multimedia resource to determine initial text content corresponding to the first multimedia resource, wherein the audio data of the first multimedia resource includes voice data of the initial text content
- S103 Determine invalid text content in the initial text content, wherein the invalid text content is text content that semantically does not express information
- S104 Determine a first playing position of the voice data of the invalid text content in the first multimedia resource
- S105 On the basis of the first playing position, cut the first multimedia resource, so as to obtain a second multimedia resource, wherein audio data of the second multimedia resource includes the voice data of the target text content and does not include the voice data of the invalid text content

[见续页]

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:

- 不包括国际检索报告, 在收到该报告后将重新公
布 (细则48.2(g))。

(57) 摘要: 摘要本公开实施例提供一种多媒体处理方法、装置、电子设备及存储介质, 通过获取第一多媒体资源; 对第一多媒体资源的音频数据进行语音识别, 确定第一多媒体资源对应的初始文本内容, 第一多媒体资源的音频数据包含初始文本内容的语音数据; 确定初始文本内容中的无效文本内容, 无效文本内容为语义上无信息表达作用的文本内容; 确定无效文本内容的语音数据在第一多媒体资源中的第一播放位置; 基于第一播放位置, 对第一多媒体资源进行裁剪, 得到第二多媒体资源, 其中, 第二多媒体资源的音频数据中包含目标文本内容的语音数据且不包含无效文本内容的语音数据。本公开实施例实现了对多媒体资源中无效内容自动剪裁, 提高了剪裁效率和剪裁效果。