

(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 2025/011651 A1

(43) 国际公布日
2025 年 1 月 16 日 (16.01.2025)

(51) 国际专利分类号:
G10L 13/10 (2013.01)

(21) 国际申请号: PCT/CN2024/105261

(22) 国际申请日: 2024 年 7 月 12 日 (12.07.2024)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202310855500.2 2023 年 7 月 12 日 (12.07.2023) CN

(71) 申请人: 支付宝 (杭州) 信息技术有限公司 (ALIPAY (HANGZHOU) INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。

(72) 发明人: 伏冠宇 (FU, Guanyu); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。杨明晖 (YANG, Minghui); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。薛吕欣 (XUE, Lvxin); 中国

浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。金春祥 (JIN, Chunxiang); 中国浙江省杭州市西湖区西溪路 556 号 8 层 B 段 801-11, Zhejiang 310000 (CN)。

(74) 代理人: 北京亿腾知识产权代理事务所 (普通合伙) (E-TONE INTELLECTUAL PROPERTY FIRM (GENERAL PARTNERSHIP)); 中国北京市海淀区中关村紫金数码园 3 号楼 707, Beijing 100190 (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: DIGITAL HUMAN MOUTH SHAPE GENERATION METHOD AND DEVICE

(54) 发明名称: 一种数字人唇形生成方法及装置

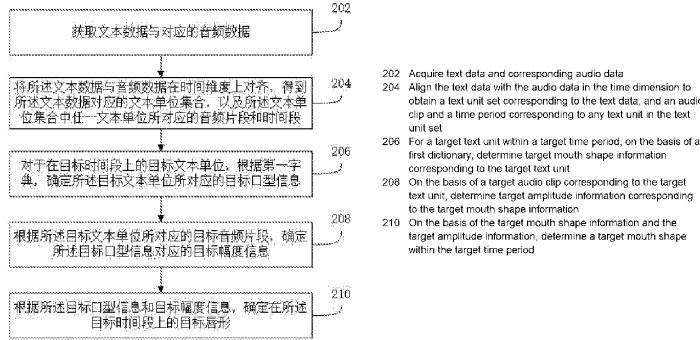


图 2

(57) Abstract: A digital human mouth shape generation method and device. According to the method, input data comprises text data and audio data corresponding to the text data; the text data and the audio data are input into an alignment module for alignment to obtain an audio clip and a time period corresponding to any text unit in the text data; then, a pre-configured text-mouth shape dictionary is used to obtain corresponding mouth shape information on the basis of the text unit; amplitude information corresponding to a mouth shape is acquired on the basis of a specific audio feature of the audio clip, and a corresponding mouth shape within one time period can be obtained by combining the mouth shape information and the amplitude information; mouth shapes within different time periods are arranged in chronological order to obtain an initial mouth shape sequence; and interpolation and smoothing operations are performed in blank time periods between adjacent mouth shapes of the mouth shape sequence to achieve the transition between different mouth shapes and obtain a final smooth mouth shape sequence.

[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: 一种数字人唇形生成方法及装置, 方法的输入数据包括文本数据和与其对应的音频数据, 将文本数据和音频数据输入到对齐模块中进行对齐, 得到文本数据中任一文本单位所对应的音频片段和时间段。然后, 使用预先配置好的文本-口型字典, 根据文本单位获得对应的口型信息; 根据音频片段的特定音频特征获取该口型对应的幅度信息, 综合口型信息和幅度信息, 便可以得到一个时间段上的对应的唇形。将不同时间段上的唇形按照时间顺序进行排列后, 得到初步的唇形序列。对这个唇形序列在相邻唇形之间的空白时间段进行插值和平滑操作, 实现不同唇形之间的过渡, 得到最终的平滑唇形序列。