

ICS 35.240.01

CCS L00

团 体 标 准

T/SCCF 16—2024

川藏沿线文旅数字化应用规范 数字人评价

Specification of digital application of culture and tourism along the sichuan tibet
highway—Digital human evaluation

2024 - 04 - 29 发布

2024 - 04 - 29 实施

四川省计算机学会 发布

目 次

前言	II
引言	III
1 范围	1
2 规范性引用文件	1
3 术语和定义	1
4 评价原则	1
4.1 公正性	1
4.2 独立性	1
4.3 操作性	1
4.4 定量定性结合	1
5 评价指标体系	1
6 评价过程	6
6.1 基本要求	6
6.2 前期准备	6
6.3 现场评价程序	6
6.4 评价结果应用	6
附录 A（规范性） 指标计算公式	7
A.1 发音准确度	7
A.2 韵律准确度	7
A.3 平均交互响应时间	7
参考文献	8

前 言

本文件按照T/CAS 1.1-2017《团体标准结构和编写指南》要求并参照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利，本文件的发布机构不承担识别专利的责任。

本文件由四川旅游学院提出。

本文件由四川省计算机学会归口。

本文件起草单位：四川旅游学院、成都产品质量检验研究院有限责任公司、元梦空间数字科技(成都)有限公司、宜宾学院、中铁二院工程集团有限责任公司、黄龙国家级风景名胜区管理局、西南交通大学、四川旅投数字信息产业发展有限责任公司、四川启明芯智能科技有限公司、四川省计算机学会科技服务中心。

本文件主要起草人：周相兵、陈云川，马洪江、张华、冉晓娟、罗晓东、陈功锁、李智、沈少朋、薛东、李晓峰、李源、王加梁、卢佳廷、温佐承、吴鹏、张智恒、郭文胜、陈潇、李兵、曾永平、郑小艳、杜晨炜、安德军、张兴博、张继、谌倩、宋昌元、罗东明。

引 言

随着3D数字人技术日益成熟并在文旅、教育、工业、政务等领域得到广泛应用，本标准旨在规范川藏沿线文化旅游区域的3D数字人技术应用，推动3D数字人与导游、讲解、主播、影视动漫、非遗等方面的结合与有效应用，以提升旅游体验，增强旅游服务的智能化和个性化，引导川藏文旅3D数字人及其相关产业健康发展。

川藏沿线文旅数字化应用规范

数字人评价

1 范围

本文件规定了川藏沿线文旅数字化应用中数字人评价原则、评价指标体系、评价过程。本文件适合于川藏沿线文旅数字化应用的数字人研发、运营。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB 21023-2007 中文语音识别系统通用技术规范
T/BAGIA 0001-2019 影视动画三维模型制作

3 术语和定义

GB/T 21024-2007、T/BAGIA 0001-2019界定的以及下列术语和定义适用于本文件。

3.1

数字人 digital human

运用数字技术创造出来的、与人类形象接近的数字化人物形象。

3.2

3D 数字人 three-dimensional digital human

具有高度、宽度和深度三个维度信息的立体形态数字人。

3.3

韵律 prosody

将各种语言学单位组织成话语或话语中关联组块的系统组织。

4 评价原则

4.1 公正性

应对客观事实情况进行公正的评审和评议。

4.2 独立性

评价组织或评价人员应独立进行评价，不应受任何其他组织和个人的干预。

4.3 操作性

评价指标、评价方法、评价程序应具有可操作性。

4.4 定量定性结合

应在定量评分基础上进行专家综合评议。

5 评价指标体系

5.1 数字人评价指标体系包括 2 项一级指标、7 项二级指标、20 项三级指标，见表 1。

5.2 数字人评价指标体系可根据数字人技术和应用场景的发展进行扩充。

5.3 开展实际评价工作时，数字人评价指标体系应根据实际评价目的和场景进行合理裁剪。

全国团体标准信息平台

表1 数字人评价指标体系

一级指标	二级指标		三级指标		
	名称	定义	名称	定义	内容描述
I1 美学指标	I1.1 角色造型	系统给出数字人身份设定、外形风格、背景、说话小动作等人物特性信息。	I1.1.1 人物特性	数字人人物身份设定、外形风格、性格等人物特性信息	<ul style="list-style-type: none"> ——数字人的身份设定情况，例如：性别、身高、体型、职业定位、现实世界与童话科幻世界定位等； ——数字人背景识别情况，例如：名字、社会关系、故事等； ——数字人外形的风格情况，例如：卡通风格、拟真风格、写实风格等； ——数字人说话时的表情、口头禅、招牌动作、透露性格的小动作等。
			I1.1.2 完好性	3D数字人形象的完好程度	<ul style="list-style-type: none"> ——是否存在严重的变形； ——是否存在严重的穿插； ——是否存在没焊接的点； ——是否存在破面； ——其他破损情况。
			I1.1.3 精细度	数字人形象的精细程度	<ul style="list-style-type: none"> ——人物模型的面数； ——人物面部细节建模情况，例如眼球、晶状体、睫毛、口腔内部结构及毛发等； ——人物面部布线结构； ——人物贴图分辨率； ——人物身体及服饰绑定层级和复杂度； ——人物身体、面部及服饰的模型点数； ——人物身体、面部及服饰的骨骼数量。
			I1.1.4 形象舒适度	数字人形象让用户生理上感到舒适的程度	<ul style="list-style-type: none"> ——用户好感度，例如：你喜欢该形象设计吗； ——自然度，例如：该形象是否自然； ——用户使用愿景，例如：你愿意使用该形象为你服务吗。

表1 数字人评价指标体系（续）

一级指标	二级指标		三级指标			
	名称	定义	名称	定义	内容描述	
I1.2 角色语音	数字人对应角色的语音情况。		I1.2.1 发音准确度	数字人从文字合成语音过程中的发音准确程度	<ul style="list-style-type: none"> ——发音不准确的表现包括漏音、吞音、多余发音、音素错误、音调错误等； ——发音准确度计算公式见附录 A 公式 (A.1) 和公式 (A.2)。 	
			I1.2.2 韵律准确度	数字人语音合成过程中的韵律准确程度	<ul style="list-style-type: none"> ——韵律包括停顿断句、音高、音长、音量、重音位置焦点位置等因素，对应了焦点发音、问句语调、感叹句语调等自然发音规律； ——韵律准确度计算公式见附录 A 公式 (A.3)。 	
			I1.2.3 语音舒适度	数字人合成语音让用户生理上感到舒适的程度	<ul style="list-style-type: none"> ——语音语调，例如：整体发音是否标准？发音吐字是否清晰？语气语调是否自然？重读发音是否得当？语速表达是否恰当； ——流畅连贯度，例如：语音表达是否流利； ——情绪饱满度，例如：按照文本语义和内容，情绪表达是否恰当； ——拟人舒适度，例如：声音拟人程度是否和真人一样？聆听该声音时，感受是否愉悦？你愿意使用该声音为你服务吗。 	
	I1.3 角色动画	生成与之相匹配的人物动画，包括面部表情和肢体动作。基于驱动方式的不同，可分为真人驱动型和智能驱动型。		I1.3.1 动作契合度	数字人动作与当下语境的契合度	<ul style="list-style-type: none"> ——嘴唇动作、眉毛与眼皮动作、眼球动作、头旋转动作； ——上身肢体动作，包括躯干关节旋转动作、大臂、小臂和手掌； ——下身肢体动作，包括大腿、小腿和脚掌、全身动作。
				I1.3.2 动作舒适度	数字人动作让用户生理上感到舒适的程度	<ul style="list-style-type: none"> ——口型匹配度，例如：口型与发音匹配吗？； ——面部表情自然度，例如：面部表情是否自然； ——肢体动作自然度，例如：肢体动作是否自然。
	I1.4 角色场景适配	数字人在不同场景的适应		I1.4.1 使用场景	虚实场景互动情况	<ul style="list-style-type: none"> ——虚拟场景与真实场景互动。

一级指标	二级指标		三级指标		
	名称	定义	名称	定义	内容描述
	度	情况。	I1.4.2 道具与互动		——动作与道具互动的流畅度、超越只能说话的范畴。
I2 技术指标	I2.1 交互处理	可使数字人具备识别感知和理解决策的能力,即通过语音语义识别等智能技术识别用户输入信息,在此基础上理解用户意图,并根据用户当前意图决定数字人后续的语音和动作,驱动人物开启新一轮交互。	I2.1.1 语音识别准确度	数字人对用户进行语音识别的性能表现	——应符合 GB/T 21023-2007 中的 5.2.1、5.2.2、5.2.3 要求。
			I2.1.2 平均交互响应时间	与用户进行一轮交互的平均响应时间	——平均交互响应时间计算公式见附录 A 公式(A.4)。
			I2.1.3 用户交互体验	用户基于对参评系统的交互体验	——交互准确度,例如:咨询问答情景下,系统是否能够准确的解答用户问题?任务执行情景下,系统是否能准确联系上下文帮助用户完成任务?闲聊情景下,系统是否能给出满意的回复?系统是否能准确识别打断; ——风格一致性,例如:在交互过程中,系统中的数字人是否保持前后风格一致,符合设定人物性格特征。
	I2.2 多模态输入	接受用户输入信息的情况。	I2.2.1 多模态输入方式	数字人系统支持的感知输入方式种类	——数字人是否支持文字、语音、图像、触控等。
	I2.3 多模态输出	将输出结果呈现给用户的情况。	I2.3.1 多模态输出方式	数字人系统支持的多模态输出方式种类	——数字人是否支持手机、电视、投影、LED 显示、裸眼立体、VR、AR 显示。
			I2.3.2 视频合成实时率	数字人视频合成实时率	——视频合成耗时与输出视频时长比值。
			I2.3.3 流畅度	数字人系统生成数字人视频的流畅度	——视频帧率,即 FPS 值(单位:帧/秒)。
			I2.3.4 画面准确率	数字人系统生成固定帧数视频时画面的准确率	——若出现跳帧、卡顿等错误均视为画面不准确。
			I2.3.5 音视频匹配度	系统生成固定时长(单位:s)视频时音视频的匹配度	——若出现口型多余、缺失,音频提前、延迟等错位均视为音视频不匹配。

6 评价过程

6.1 基本要求

数字人评价应按以下要求：

- a) 应根据评价需求成立评价小组；
- b) 评价小组成员应包含数字人应用领域专业人员；
- c) 评价小组成员人数应为奇数，人数应不少于 5 人；
- d) 评价周期和评价时间应根据需求确定。

6.2 前期准备

前期准备主要工作包括：成立评价小组、制定评价工作方案、收集项目基础资料、形成评价指标体系、制定专家评分表及其相关文本等，评价小组宜对表1进行裁剪形成实际的评价指标体系。

6.3 现场评价程序

评价组应当依据评价方案，对评价对象按照下列程序进行：

- a) 宣布评价组成员、评价程序及有关事宜；
- b) 听取评价对象单位工作汇报；
- c) 查阅相应的文件、记录、标准文本等资料；
- d) 考核现场；
- e) 随机调查用户满意程度；
- f) 对照评价计分表进行测评；
- g) 形成评价结论；
- h) 向评价对象通报评价情况，提出改进意见和建议。

6.4 评价结果应用

评价结果可作为数字人开发运营组织改进产品和服务的重要依据。

附 录 A
(规范性)
指标计算公式

A.1 发音准确度

发音准确度可采用发音字准确度或发音句准确度，其分别的计算公式见公式 (A.1)、(A.2)。

$$R_{wc} = \left(1 - \frac{N_{ew}}{N_w}\right) \times 100\% \dots\dots\dots (A.1)$$

式中：

N_w —文本总字数，单位为个；

N_{ew} —发音错误字数（多个发音错误字数之和），单位为个；

R_{wc} —发音字准确率。

$$R_{sc} = \left(1 - \frac{N_{es}}{N_s}\right) \times 100\% \dots\dots\dots (A.2)$$

式中：

N_s —文本总字数，单位为个；

N_{es} —发音错误句数，单位为个；

R_{sc} —发音句字准确率。

A.2 韵律准确度

韵律准确度计算公式见公式 (A.3)。

$$R_{pc} = \frac{N_{pc}}{N} \times 100\% \dots\dots\dots (A.3)$$

式中：

N_{pc} —停顿正确用例数，单位为个；

N —总用例数，单位为个；

R_{pc} —韵律准确率。

A.3 平均交互响应时间

平均交互响应时间计算公式见公式 (A.4)。

$$T = \frac{\sum_{i=1}^N (Ts_i - Te_i)}{N} \dots\dots\dots (A.4)$$

式中：

i —表示第*i*轮交互；

Ts_i —系统开始反馈的时间点，单位为秒 (s)；

Te_i —用户输入信息结束的时间点，单位为秒 (s)；

N —测试总次数；

T —平均交互响应时间，单位为秒 (s)。

参 考 文 献

- [1] 产品言语. 3D数字人制作标准. <https://zhuanlan.zhihu.com/p/663971599>
- [2] 中国通信标准化协会. 虚拟数字人指标要求和评估方法 第1部分: 参考框架. <https://www.doc88.com/p-50659701571738.html>
-