

ICS 33.160

CCS M 61

# 团体标准

T/TAF 203—2024

## 面向 3D 身份型虚拟数字人技术要求和测试方法

Technical requirements and test methods of digital human for 3D  
identity type oriented

2024-02-23 发布

2024-02-23 实施

电信终端产业协会 发布



## 目 次

前言 .....	II
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 缩略语 .....	2
5 技术要求 .....	2
5.1 载体要求 .....	2
5.2 整体设定要求 .....	2
5.3 静态形象要求 .....	3
5.4 动态行为要求 .....	4
5.5 语音能力要求 .....	5
5.6 交互能力要求 .....	6
5.7 易用性要求 .....	6
6 测试方法 .....	7
6.1 载体测试方法 .....	7
6.2 整体设定测试方法 .....	7
6.3 静态形象测试方法 .....	8
6.4 动态行为测试方法 .....	9
6.5 语音能力测试方法 .....	11
6.6 交互能力测试方法 .....	13
6.7 易用性测试方法 .....	15
参考文献 .....	18

## 前 言

本文件按照 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。

本文件中的某些内容可能涉及专利。本标准的发布机构不承担识别专利的责任。

本文件由电信终端产业协会（TAF）提出并归口。

本文件起草单位：中国信息通信研究院、苏州跬步信息技术有限公司、北京元宇宙文化有限公司、深圳信息通信研究院、阿里巴巴（中国）有限公司。

本文件主要起草人：马蓁蓁、郭隆庆、刘轶、赵威、边元松、齐萱、王国林、姚金冶、周玉国、汪辰、李朋、段虎才、王红、沈煜。



# 面向 3D 身份型虚拟数字人技术要求和测试方法

## 1 范围

本文件规定了 3D 身份型虚拟数字人的技术要求和测试方法，主要描述了载体、整体设定、静态形象、动态行为、语音能力、交互能力、易用性等方面技术要求的测试方法。

本文件适用于各种制式的 3D 身份型虚拟数字人。

## 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/T 21024—2007 中文语音合成系统通用技术规范

T/BAGIA 0001—2019 影视动画三维模型制作

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**虚拟数字人** digital human

简称数字人或虚拟人，是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

**身份型虚拟数字人** identity digital human

具有身份性的虚拟数字人，多以虚拟 IP 或偶像呈现，能够为未来的虚拟化世界提供人的核心交互中介。

### 3.3

**模型** model

按照对象所需呈现的结构或动势等，通过专用软件完成对象每个表面拓扑结构的数学表示，从而在虚拟三维空间中塑造出的物体。

[来源：T/BAGIA 0001—2019 ]

### 3.4

**韵律** prosody

一般也称为超音段特征，是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。从物理角度，指基频、时长和强度等声学参数；从语言学的角度，指音段层次以上的音系组织。

[来源：GB/T 21024—2007 ]

## 4 缩略语

下列缩略语使用于本文件。

3D: 三维 (3-Dimensional)

3S 贴图: 磁表面散射贴图 (Sub-Surface Scattering)

A0 贴图: 环境遮挡贴图 (Ambient Occlusion)

AR: 增强现实技术 (Virtual Reality)

FPS: 每秒传输帧数 (Frames Per Second)

LED: 发光二极管 (light-emitting diode)

PX: 像素 (Pixel)

VR: 虚拟现实技术 (Virtual Reality)

## 5 技术要求

### 5.1 载体要求

#### 5.1.1 基础载体

基础载体要求如下:

- a) 具备在手机、电视、投影、LED 大屏等任意平台中运行;
- b) 具备在裸眼立体、VR、AR 显示等任意环境里直接展示运行。

#### 5.1.2 兼容载体

兼容载体要求如下:

- a) 具备在相同类型的平台中, 多品牌、多型号设备载体上应能正常运行;
- b) 具备在相同类型的环境中, 多品牌、多型号设备载体上应能正常运行。

### 5.2 整体设定要求

#### 5.2.1 身份设定

身份设定要求如下:

- a) 人物姓名、性别、身高、体型属性, 应与设计匹配;
- b) 人物特点、特长 (技能) 属性, 应与设计匹配;
- c) 人物与现实世界定位, 应和设计匹配。

#### 5.2.2 外形设定

人物外形, 应和设计匹配。

#### 5.2.3 建模风格

人物模型应属于卡通、拟人、写实、超写实四种风格。

### 5.3 静态形象要求

#### 5.3.1 规范性

规范性要求如下：

- a) 人物不应有破面问题；
- b) 人物不应有严重穿插问题；
- c) 人物不应有严重的变形问题；
- d) 人物不应有材质球渲染问题；
- e) 人物不应有其他破损情况问题；
- f) 人物不应有没焊接的点问题；
- g) 人物不应有阴影问题；
- h) 人物不应有时装穿模问题；
- i) 人物不应有碰撞问题；
- j) 人物不应有法线问题；
- k) 人物不应有口型问题；
- l) 人物不应有资源丢失问题；
- m) 人物不应有骨骼异常问题。

#### 5.3.2 精细度

精细度要求如下：

- a) 人物面数，不低于 3000 面；
- b) 人物贴图数量，不低于 4 张；
- c) 人物贴图分辨率，不低于 1024px\*1024px；
- d) 人物毛发贴图精度，不低于 2048px\*2048px；
- e) 人物应使用 A0 贴图；
- f) 人物眼睛模型应有颜色贴图；
- g) 人物眼睛模型应有粗糙贴图；
- h) 人物皮肤应有 3S 贴图；
- i) 人物应有眼球结构；
- j) 人物应有眼球晶状体结构；
- k) 人物应有口腔内部结构；
- l) 人物模型应为可编辑多边形；
- m) 人物骨骼应是 Biped 骨骼；
- n) 人物骨骼点应在模型骨骼关节处；
- o) 人物角色脖子骨骼应带有角度；
- p) 人物发型帽饰动态骨骼数量应为 5 根以上；
- q) 人物所有骨骼应至少 2Link 以上的权重；
- r) 人物蒙皮后角色 A-pose 应改为标准 T-pose；
- s) 人物转化为 T-pose 的角色蒙皮不应出现错误；
- t) 人物布线走向应和骨骼方向垂直。

#### 5.3.3 可编辑性

可编辑性要求如下：

- a) 人物应可调整脸型;
- b) 人物应可调整五官;
- c) 人物应可调整妆容;
- d) 人物应可调整肤色;
- e) 人物应可调整身高;
- f) 人物应可调整体重;
- g) 人物应可更换衣服;
- h) 人物应可更换发型;
- i) 人物应可更换装饰品。

## 5.4 动态行为要求

### 5.4.1 动作契合度

#### 5.4.1.1 嘴唇动作

嘴唇动作要求如下:

- a) 人物按字发音时, 嘴形和发音应同步;
- b) 人物按句发音时, 前后若干相邻发音和嘴形应同步;
- c) 人物按照快乐的文字发音时, 嘴型应嘴唇上拉、嘴角收紧;
- d) 人物按照生气的文字发音时, 嘴型应下唇上推、撅嘴;
- e) 人物按照不同意见文字发音时, 嘴型应撅嘴。

#### 5.4.1.2 眉毛与眼皮动作

眉毛与眼皮动作要求如下:

- a) 人物说话时, 眉毛应根据表达的语意动态变化、眼皮应眨眼;
- b) 人物表达快乐时, 内眉毛应上扬、上眼睑应上扬;
- c) 人物表达生气时, 内眉毛应下降、脸颊应提升并眼轮外圈收紧;
- d) 人物表达不同意见时, 应皱眉、眼轮内圈收紧;
- e) 人物应可以闭眼、眨眼。

#### 5.4.1.3 眼球动作

眼球动作要求如下:

- a) 人物眼球应可以旋转;
- b) 人物表达快乐时, 瞳孔应会放大;
- c) 人物表达生气时, 瞳孔应会缩小;
- d) 人物表达不同意见时, 应眯着眼睛;
- e) 人物表达惊讶、好奇时, 应睁大眼睛。

#### 5.4.1.4 头旋转动作

头旋转动作要求如下:

- a) 人物说话时, 应旋转到对应位置, 并且根据表达的语意动态变化;
- b) 人物表达相同意见时, 头应点头;
- c) 人物表达不同意见时, 头应摇头;



- d) 人物头应可以左转、右转、向上、向下、左倾斜、右倾斜、前倾、后仰；
- e) 人物头应可以上下摆动、左右摆动、上扬再左/右倾。

#### 5.4.1.5 上身肢体动作

上身肢体动作要求如下：

- a) 人物说话时，上身肢体动作应根据表达的语意动态变化；
- b) 人物上身肢体应可以按照实体的属性或行为动作；
- c) 人物上身肢体应可以按照抽象性的概念动作；
- d) 人物上身肢体应可以模拟真实人类进行相关生产生活动作。

#### 5.4.1.6 下身肢体动作

下身肢体动作要求如下：

- a) 人物下身肢体应可以模拟真人自然地维持身体的平衡；
- b) 人物下身肢体应可以走、跑、跳实现身体位移；
- c) 人物进行多种行动方式动作，动作应自然切换；
- d) 人物内心着急状态时，走路/跑步姿态和频率与普通应不一致。

#### 5.4.1.7 全身动作

根据人物特性，执行相关场景动作，人物全身动作应协调一致。

#### 5.4.2 实时驱动

实时驱动要求如下：

- a) 人物在手机及摄像头面捕驱动下，面部表情应具有准确度、即时性（时延）、平滑度（抖动）；
- b) 人物在手臂传感系统、手套等设备及半身摄像头设备下，上肢动作应具有准确度、即时性（时延）、平滑度（抖动）、协调性；
- c) 人物在光学系统、惯性系统等设备及摄像头动捕设备下，全身动作应具有准确度、即时性（时延）、平滑度（抖动）、协调性。

#### 5.4.3 交互准确性

交互准确性要求如下：

- a) 两个人物进行交互，表情应契合；
- b) 两个人物进行交互，动作应契合。

### 5.5 语音能力要求

#### 5.5.1 发音准确度

发音准确度要求如下：

- a) 人物按字发音，准确率不低于 95%；
- b) 人物按句发音。准确率不低于 95%。

#### 5.5.2 韵律准确度

人物按段停顿断句，韵律准确率不低于 95%。

### 5.5.3 语音效果

语音效果要求如下：

- a) 人物应可以使用不同音色声音说话；
- b) 人物应可以调整说话速度。

### 5.5.4 语言能力

语言能力要求如下：

- a) 人物应可以使用不同方言说话；
- b) 人物应可以使用不同外语说话。

### 5.5.5 交互准确度

两个人物进行语音交互，语音识别准确率超过 95%。

## 5.6 交互能力要求

### 5.6.1 输入

输入要求如下：

- a) 人物应可以点触控制；
- b) 人物应可以图像控制；
- c) 人物应可以文字控制；
- d) 人物应可以语音控制；
- e) 人物应可以动捕硬件控制；
- f) 人物应可以面捕硬件控制。

### 5.6.2 处理

处理要求如下：

- a) 人物与用户单轮对话，意图识别率不低于 95%；
- b) 人物与用户多轮对话，任务完成率不低于 95%；
- c) 人物与用户多种语言对话，意图识别率不低于 95%；
- d) 人物与用户交互，平均响应时间不低于 2 秒；
- e) 人物与用户长时间交互，不应出现卡顿、反应慢、延迟等现象；
- f) 用户使用生气或者愤怒等情绪激动的词语与人物交互，系统应识别并有效的给予安抚；
- g) 用户使用欺诈、恶意自扰、自杀、恢复设置等业务风险关键词交互，系统应能识别和规避。

### 5.6.3 输出

输出要求如下：

- a) 系统生成数字人视频的流畅度，视频帧率不低于 30FPS；
- b) 系统生成固定帧数视频，不应出现跳帧、卡顿等错误；
- c) 固定时长（单位：s）视频，不应出现口型多余、缺失，音频提前、延迟等错误。

## 5.7 易用性要求

### 5.7.1 形象舒适性

从好感度、自然度、契合度、使用愿景四个方面进行评分。

### 5.7.2 语音舒适性

从语音语调、流畅连贯度、情绪饱满度、拟人舒适度四个方面进行评分。

### 5.7.3 动画舒适性

从口型匹配度、面部表情自然度、肢体动作自然度、动作衔接连贯性四个方面进行评分。

### 5.7.4 用户交互体验

从交互准确度、风格一致性两个方面进行评分。

## 6 测试方法

### 6.1 载体测试方法

#### 6.1.1 基础载体

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在多个平台和多个环境里直接展示运行，按照 5.1.1 基础载体中规定的技术要求内容，检查虚拟数字人运行是否正常。
- b) 预期结果：
  - 1) 3D 身份型虚拟数字人能够在任意一个平台中正常运行；
  - 2) 3D 身份型虚拟数字人能够在任意一个环境中正常展示运行。

#### 6.1.2 兼容载体

测试方法如下：

- a) 测试方法：
  - 1) 将 3D 身份型虚拟数字人，在多于十台相同类型平台终端上运行，按照 5.1.2 兼容载体中规定的技术要求内容，检查虚拟数字人运行是否正常；
  - 2) 将 3D 身份型虚拟数字人，在多于十台相同类型环境终端上运行，按照 5.1.2 兼容载体中规定的技术要求内容，检查虚拟数字人运行是否正常。
- b) 预期结果：
  - 1) 3D 身份型虚拟数字人能够在相同类型的平台中，多品牌、多型号设备载体上正常运行；
  - 2) 3D 身份型虚拟数字人能够在相同类型的环境中，多品牌、多型号设备载体上正常运行。

### 6.2 整体设定测试方法

#### 6.2.1 身份设定

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，按照 5.2.1 身份设定中规定的技术要求内容，检查虚拟数字人人物各种属性和定位是否和设计匹配。
- b) 预期结果：
  - 1) 人物姓名、性别、身高、体型属性和设计匹配；
  - 2) 人物特点、特长（技能）属性和设计匹配；

- 3) 人物与现实世界定位和设计匹配。

### 6.2.2 外形设定

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，按照 5.2.2 外形设定中规定的技术要求内容，检查虚拟数字人人物外形是否和设计匹配。
- b) 预期结果：人物外形和设计匹配。

### 6.2.3 建模风格

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，按照 5.2.3 建模风格中规定的技术要求内容，检查虚拟数字人人物模型属于哪一种风格。
- b) 预期结果：人物模型属于卡通、拟人、写实、超写实四种其中一种风格。

## 6.3 静态形象测试方法

### 6.3.1 规范性

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，按照 5.3.1 规范性中规定的技术要求内容，检查虚拟数字人是否存在规范性问题。
- b) 预期结果：
  - 1) 人物不存在破面；
  - 2) 人物不存在严重穿插；
  - 3) 人物不存在严重的变形；
  - 4) 人物不存在材质球渲染问题；
  - 5) 人物不存在其他破损情况；
  - 6) 人物不存在没焊接的点；
  - 7) 人物不存在阴影问题；
  - 8) 人物不存在时装穿模问题；
  - 9) 人物不存在碰撞问题；
  - 10) 人物不存在法线问题；
  - 11) 人物不存在口型问题；
  - 12) 人物不存在资源丢失问题；
  - 13) 人物不存在骨骼异常问题。

### 6.3.2 精细度

测试方法如下：

- a) 测试方法：在专业的虚拟数字人开发工具上，按照 5.3.2 精细度中规定的技术要求内容，检查虚拟数字人精细度是否符合要求。
- b) 预期结果：
  - 1) 卡通人物模型面数超过 3000 面，拟人人物面数超过 1 万面，写实人物面数超过 5 万面，超写实人物面数超过 500 万面；
  - 2) 人物贴图数量超过 4 张；

- 3) 人物贴图分辨率高于 1024px\*1024px;
- 4) 人物毛发贴图精度高于 2048px\*2048px;
- 5) 人物有使用 AO 贴图;
- 6) 人物眼睛模型有颜色贴图;
- 7) 人物眼睛模型有粗糙贴图;
- 8) 人物皮肤有 3S 贴图;
- 9) 人物有眼球结构;
- 10) 人物有眼球晶状体结构;
- 11) 人物有口腔内部结构;
- 12) 人物模型为可编辑多边形;
- 13) 人物骨骼是 Biped 骨骼;
- 14) 人物骨骼点在模型骨骼关节处;
- 15) 人物角色脖子骨骼带有角度;
- 16) 人物发型帽饰动态骨骼数量为 5 根以上;
- 17) 人物所有骨骼至少 2Link 以上的权重;
- 18) 人物蒙皮后角色 A-pose 改为标准 T-pose;
- 19) 人物转化为 T-pose 的角色蒙皮没有错误;
- 20) 人物布线走向和骨骼方向垂直。

### 6.3.3 可编辑性

测试方法如下:

- a) 测试方法: 将 3D 身份型虚拟数字人, 在可适配载体终端上运行, 按照 5.3.3 可编辑性中规定的技术要求内容, 检查虚拟数字人是否可以编辑。
- b) 预期结果:
  - 1) 人物可调整脸型;
  - 2) 人物可调整五官;
  - 3) 人物可调整妆容;
  - 4) 人物可调整肤色;
  - 5) 人物可调整身高;
  - 6) 人物可调整体重;
  - 7) 人物可更换衣服;
  - 8) 人物可更换发型;
  - 9) 人物可更换装饰品。

## 6.4 动态行为测试方法

### 6.4.1 动作契合度

#### 6.4.1.1 嘴唇动作

测试方法如下:

- a) 测试方法: 将 3D 身份型虚拟数字人, 在可适配载体终端上运行, 按照 5.4.1.1 嘴唇动作中规定的技术要求内容, 检查虚拟数字人嘴唇动作是否符合要求。
- b) 预期结果:

- 1) 人物按字发音时，嘴形和发音同步；
- 2) 人物按句发音时，前后若干相邻发音和嘴形同步；
- 3) 人物按照快乐的文字发音时，嘴唇上拉、嘴角收紧；
- 4) 人物按照生气的文字发音时，下唇上推、撅嘴；
- 5) 人物按照不同意见文字发音时，撅嘴。

#### 6.4.1.2 眉毛与眼皮动作

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，按照 5.4.1.2 眉毛与眼皮动作中规定的技术要求内容，检查虚拟数字人眉毛与眼皮动作是否符合要求。
- b) 预期结果：
  - 1) 人物说话时，眉毛根据表达的语意动态变化、眼皮眨眼；
  - 2) 人物表达快乐时，内眉毛上扬、上眼睑上扬；
  - 3) 人物表达生气时，内眉毛下降、脸颊提升并眼轮外圈收紧；
  - 4) 人物表达不同意见时，皱眉、眼轮内圈收紧；
  - 5) 人物可以闭眼、眨眼。

#### 6.4.1.3 眼球动作

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，按照 5.4.1.3 眼球动作中规定的技术要求内容，检查虚拟数字人眼球动作是否符合要求。
- b) 预期结果：
  - 1) 人物眼球可以旋转；
  - 2) 人物表达快乐时，瞳孔会放大；
  - 3) 人物表达生气时，瞳孔会缩小；
  - 4) 人物表达不同意见时，眯着眼睛；
  - 5) 人物表达惊讶、好奇时，睁大眼睛。

#### 6.4.1.4 头旋转动作

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，按照 5.4.1.4 头旋转动作中规定的技术要求内容，检查虚拟数字人头旋转动作是否符合要求。
- b) 预期结果：
  - 1) 人物说话时，头旋转到对应位置，并且根据表达的语意动态变化；
  - 2) 人物表达相同意见时，点头；
  - 3) 人物表达不同意见时，摇头；
  - 4) 人物头可以左转、右转、向上、向下、左倾斜、右倾斜、前倾、后仰；
  - 5) 人物头可以上下摆动、左右摆动、上扬再左/右倾。

#### 6.4.1.5 上身肢体动作

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，按照 5.4.1.5 上身肢体动作中规定的技术要求内容，检查虚拟数字人上身肢体动作是否符合要求。

- b) 预期结果:
- 1) 人物说话时, 上身肢体动作根据表达的语意动态变化;
  - 2) 人物上身肢体可以按照实体的属性或行为动作;
  - 3) 人物上身肢体可以按照抽象性的概念动作;
  - 4) 人物上身肢体可以模拟真实人类进行相关生产生活动作。

#### 6.4.1.6 下身肢体动作

测试方法如下:

- a) 测试方法: 将 3D 身份型虚拟数字人, 在可适配载体终端上运行, 按照 5.4.1.6 下身肢体动作中规定的技术要求内容, 检查虚拟数字人下身肢体动作是否符合要求。
- b) 预期结果:
- 1) 人物下身肢体可以模拟真人自然地维持身体的平衡;
  - 2) 人物下身肢体可以走、跑、跳实现身体位移;
  - 3) 人物进行多种行动方式动作, 动作可以自然切换;
  - 4) 人物内心着急状态时, 走路/跑步姿态和频率与普通一致。

#### 6.4.1.7 全身动作

测试方法如下:

- a) 测试方法: 将 3D 身份型虚拟数字人, 在可适配载体终端上运行, 按照 5.4.1.7 全身动作中规定的技术要求内容, 检查虚拟数字人全身肢体动作是否符合要求。
- b) 预期结果: 人物全身动作协调一致。

#### 6.4.2 实时驱动

测试方法如下:

- a) 测试方法: 将 3D 身份型虚拟数字人, 在多种可适配载体终端上运行, 按照 5.4.2 实时驱动中规定的技术要求内容, 检查虚拟数字人实时驱动是否符合要求。
- b) 预期结果:
- 1) 人物在手机及摄像头面捕驱动下, 面部表情准确、延迟低、无抖动;
  - 2) 人物在手臂传感系统、手套等设备及半身摄像头设备下, 上肢动作表情准确、延迟低、无抖动;
  - 3) 人物在光学系统、惯性系统等设备及摄像头动捕设备下, 全身动作表情准确、延迟低、无抖动。

#### 6.4.3 交互准确性

测试方法如下:

- a) 测试方法: 将 3D 身份型虚拟数字人, 在可适配载体终端上运行, 按照 5.4.3 交互准确性中规定的技术要求内容, 检查虚拟数字人交互准确性是否符合要求。
- b) 预期结果:
- 1) 两个人物进行交互表情契合;
  - 2) 两个人物进行交互动作契合。

#### 6.5 语音能力测试方法



### 6.5.1 发音准确度

测试方法如下：

a) 测试方法：

- 1) 将 3D 身份型虚拟数字人，在可适配载体终端上运行；
- 2) 构建包含普通文本和特殊文本的文本测试集，生成对应视频，统计发音错误的字符数，以及发音错误的句数，按照 5.5.1 发音准确度中规定的技术要求内容，检查虚拟数字人发音准确度是否符合要求；
- 3) 发音字准确率，计算方法如公式 1 所示：

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

式中：

- $N_w$ ——文本总字数，单位为个；
- $N_{ew}$ ——发音错误字数（多种发音错误字数之和），单位为个；
- $R_{wc}$ ——发音字准确率。

4) 发音句准确率，计算方法如公式 2 所示：

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

式中：

- $N_s$ ——文本总句数，单位为个；
- $N_{es}$ ——发音错误句数，单位为个；
- $R_{sc}$ ——发音句准确率。

b) 预期结果：

- 1) 人物按字发音字准确率超过 95%；
- 2) 人物按句发音句准确率超过 95%。

### 6.5.2 韵律准确度

测试方法如下：

a) 测试方法：

- 1) 将 3D 身份型虚拟数字人，在可适配载体终端上运行；
- 2) 使用普通文本测试数据，输入一段文本，生成对应视频，记录文字停顿、断句位置与标准答案比对，一个句子为一个用例，统计停顿正确用例数，按照 5.5.2 韵律准确度中规定的技术要求内容，检查虚拟数字人韵律准确度是否符合要求；
- 3) 韵律准确率，计算方法如公式 3 所示：

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

式中：

- $N_{pc}$ ——停顿正确用例数，单位为个；



$N$  ——总用例数，单位为个；

$R_{oc}$  ——韵律准确率。

- b) 预期结果：人物按段停顿断句韵律准确率超过 95%。

### 6.5.3 语音效果

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，使用不同音色和说话速度，按照 5.5.3 语音效果中规定的技术要求内容，检查虚拟数字人语音效果是否符合要求。
- b) 预期结果：
- 1) 人物可以使用不同音色声音说话；
  - 2) 人物可以调整说话速度。

### 6.5.4 语言能力

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，使用不同方言和外语，按照 5.5.4 语言能力中规定的技术要求内容，检查虚拟数字人语言能力是否符合要求。
- b) 预期结果：
- 1) 人物可以使用不同方言说话，包括东北话、四川话、粤语等；
  - 2) 人物可以使用不同外语说话，包括英语、法语、俄语等。

### 6.5.5 交互准确度

测试方法如下：

- a) 测试方法：
- 1) 将 3D 身份型虚拟数字人，在可适配载体终端上运行，两个虚拟数字人进行语音交互，按照 5.5.5 交互准确性中规定的技术要求内容，检查虚拟数字人交互准确度是否符合要求；
  - 2) 语音识别准确率，按照 GB/T 21024—2007 中的 5.2.1、5.2.2、5.2.3 进行评估。
- b) 预期结果：两个人物进行语音交互，语音识别准确率超过 95%。

## 6.6 交互能力测试方法

### 6.6.1 输入

测试方法如下：

- a) 测试方法：将 3D 身份型虚拟数字人，在多种可适配载体终端上，通过不同方式控制虚拟数字人，按照 5.6.1 输入能力中规定的技术要求内容，检查虚拟数字输入能力是否符合要求。
- b) 预期结果：
- 1) 人物可以点触控制；
  - 2) 人物可以图像控制；
  - 3) 人物可以文字控制；
  - 4) 人物可以语音控制；
  - 5) 人物可以动捕硬件控制；
  - 6) 人物可以面捕硬件控制。

### 6.6.2 处理

测试方法如下：

a) 测试方法：

- 1) 将 3D 身份型虚拟数字人，在可适配载体终端上，通过不同语音对话方式、不同语言与虚拟数字人交互，按照 5.6.2 处理能力中规定的技术要求内容，检查虚拟数字人处理能力是否符合要求。
- 2) 单轮对话意图识别率，计算方法如公式 4 所示：

$$R_{arc} = \frac{N_{ar}}{N} \times 100\% \dots\dots\dots (4)$$

式中：

- $N_{ar}$ ——精准识别的数量，单位为个；
- $N$ ——单轮对话测试总数，单位为个；
- $R_{arc}$ ——意图识别率。

- 3) 多轮对话任务完成率，计算方法如公式 5 所示：

$$R_{tc} = \frac{L_1}{N} \times 100\% \dots\dots\dots (5)$$

式中：

- $L_1$ ——任务完成量（多轮对话每次测试所有轮次均顺利识别视为任务完成），单位为个；
- $N$ ——多轮对话测试总数，单位为个；
- $R_{tc}$ ——任务完成率。

- 4) 多种语言对话意图识别率，计算方法如公式6所示：

$$R_{cr} = \frac{L_2}{M} \times 100\% \dots\dots\dots (6)$$

式中：

- $L_2$ ——成功识别的总轮次，单位为个；
- $M$ ——测试总轮次数；
- $R_{cr}$ ——下文识别率。

- 5) 交互的平均响应时间，计算方法如公式 7 所示：

$$\bar{T} = \frac{\sum_{i=1}^N (Ts_i - Te_i)}{N} \dots\dots\dots (7)$$

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；
- $\bar{T}$ ——平均交互响应时间，单位为秒（s）。

b) 预期结果：

- 1) 人物单轮对话意图识别率超过 95%；
- 2) 人物多轮对话任务完成率超过 95%；

- 3) 人物多种语言对话意图识别率 95%;
- 4) 人物交互的平均响应时间不超过 2 秒;
- 5) 人物长时间交互, 不会出现卡顿、反应慢、延迟等现象;
- 6) 人物能够识别用户情绪, 并有效的给予安抚;
- 7) 人物能够识别业务风险, 并有效进行规避。

### 6.6.3 输出

测试方法如下:

- a) 测试方法: 将 3D 身份型虚拟数字人, 生成视频输出, 按照 5.6.3 输出能力中规定的技术要求内容, 检查虚拟数字人输出能力是否符合要求。
- b) 预期结果:
  - 1) FPS 值超过 30;
  - 2) 视频画面不会出现跳帧、卡顿等错误;
  - 3) 视频画面不会出现口型多余、缺失, 音频提前、延迟等错误。

## 6.7 易用性测试方法

### 6.7.1 形象舒适性

测试方法如下:

- a) 测试方法: 将 3D 身份型虚拟数字人, 在可适配载体终端上运行, 根据用户看到的数字人形象质量, 参照表 1 的李克特量表从形象舒适性主观评分规则进行主观打分来评价质量优劣, 1 最差~5 最优, 最后取均值。

表1 形象舒适性主观评分规则

测试内容	描述	5	4	3	2	1
好感度	你喜欢该形象的设计吗?	十分喜欢	比较喜欢	一般	不太喜欢	十分不喜欢
自然度	该形象是否自然?	十分自然	比较自然	基本自然	不太自然	十分不自然
契合度	该形象是否和设计或实体契合?	十分契合	比较契合	基本契合	不太契合	十分不契合
使用愿景	你愿意使用该形象为你服务吗?	非常愿意	比较愿意	一般	不太愿意	十分不愿意

- b) 预期结果: 从好感度、自然度、契合度、使用愿景四个方面分别给出一个主观评分。

### 6.7.2 语音舒适性

测试方法如下:

- a) 测试方法: 将 3D 身份型虚拟数字人, 在可适配载体终端上运行, 根据用户听到的数字人声音质量, 参照表 2 的李克特量表从语音舒适性主观评分规则进行主观打分来评价质量优劣, 1 最差~5 最优, 最后取均值。

表2 语音舒适性主观评分规则

测试内容	描述	5	4	3	2	1
语音语调	整体发音是否标准?	十分标准	比较标准	基本标准	个别标准	十分不标准
	发音吐字是否清晰?	十分清晰	比较清晰	基本清晰	不太清晰	十分不清晰

表2 语音舒适性主观评分规则（续）

测试内容	描述	5	4	3	2	1
语音语调	断词断句、停顿是否恰当？	十分恰当	比较恰当	基本恰当	不太恰当	很不恰当
	语气语调是否自然？	十分自然	比较自然	基本自然	不太自然	十分不自然
	重读发音是否得当？	十分恰当	比较恰当	基本恰当	不太恰当	很不恰当
	语速表达是否恰当？	十分恰当	比较恰当	基本恰当	不太恰当	很不恰当
流畅连贯度	语音表达是否流利？	十分自然	比较自然	基本自然	不太自然	十分不自然
情绪饱满度	按照文本语义和内容，情绪表达是否恰当？	十分恰当	比较恰当	基本恰当	不太恰当	很不恰当
拟人舒适度	声音拟人程度是否和真人一样？	完全无法区分	比较相似，与真人语音有细微区别	基本相似	不太一样	完全不一样
	聆听该声音时，感受是否愉悦？	十分愉悦	比较愉悦	一般	不太愉悦	十分不愉悦
	你愿意使用该声音为你服务吗？	十分愿意	比较愿意	一般	不太愿意	十分不愿意

b) 预期结果：从语音语调、流畅连贯度、情绪饱满度、拟人舒适度四个方面分别给出一个主观评分。

### 6.7.3 动画舒适性

测试方法如下：

a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，根据用户看到的数字人动画效果，参照表 3 的李克特量表从动画舒适性主观评分规则进行主观打分来评价质量优劣，1 最差~5 最优，最后取均值。

表3 动画舒适性主观评分规则

测试内容	描述	5	4	3	2	1
口型匹配度	口型与发音匹配吗？	完全匹配	比较匹配	基本匹配	不太匹配	完全不匹配
面部表情自然度	面部表情是否自然？	十分自然	比较自然	基本自然	不太自然	十分不自然
肢体动作自然度	肢体动作是否自然？	十分自然	比较自然	基本自然	不太自然	十分不自然
动作衔接连贯性	动作衔接是否连贯？	十分连贯	比较连贯	基本连贯	不太连贯	十分不连贯

b) 预期结果：从口型匹配度、面部表情自然度、肢体动作自然度、动作衔接连贯性四个方面分别给出一个主观评分。

### 6.7.4 用户交互体验

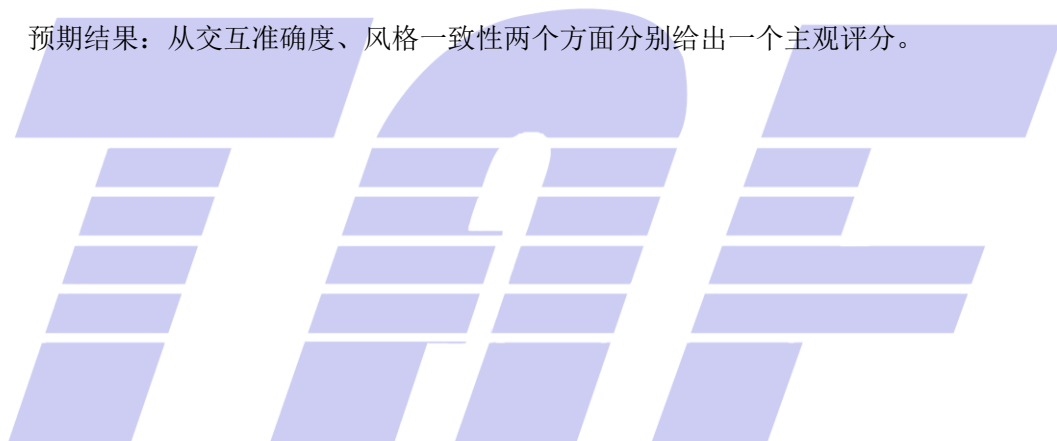
测试方法如下：

a) 测试方法：将 3D 身份型虚拟数字人，在可适配载体终端上运行，根据用户对参评数字人系统的交互体验，参照表 4 的李克特量表从用户交互体验主观评分规则进行主观打分来评价质量优劣，1 最差~5 最优，最后取均值。

表4 用户交互体验主观评分规则

测试内容	描述	5	4	3	2	1
交互准确度	咨询问答情景下，系统是否能够准确的解答用户问题？	完全准确	比较准确	基本准确	不太准确	完全不准确
	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？	完全准确	比较准确	基本准确	不太准确	完全不准确
	闲聊情景下，系统是否能给出满意的回复？	完全满意	比较满意	基本满意	不太满意	完全不满意
	系统是否能准确识别打断？	完全准确	比较准确	基本准确	不太准确	完全不准确
风格一致性	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？	十分一致	比较一致	基本一致	不太一致	完全不一致

b) 预期结果：从交互准确度、风格一致性两个方面分别给出一个主观评分。



参 考 文 献

- [1] GB/T 21024—2007 中文语音合成系统通用技术规范  
[2] T/BAGIA 0001—2019 影视动画三维模型制作
- 



电信终端产业协会团体标准  
面向 3D 身份型虚拟数字人技术要求和测试方法

T/TAF 203—2024

\*

版权所有 侵权必究

电信终端产业协会印发  
地址：北京市西城区新街口外大街 28 号  
电话：010-82052809  
电子版发行网址：[www.taf.org.cn](http://www.taf.org.cn)