

Q/027CJSC003

· · · · · 企 业 标 准

Q/027CJSC003—2023

长江证券

实时智能音视频技术规范

The enterprise standard of intelligent RTC service of
ChangJiang Securities

2023- 08 发布

2023- 08 实施

长江证券股份有限公司发布

目 次

目录

目 次..... I

前 言..... III

引 言..... IV

长江证券实时智能音视频技术规范..... 1

1 范围..... 1

2 规范性引用文件..... 1

3 定义和术语..... 1

 3.1 人工智能 Artificial Intelligent..... 1

 3.2 机器人流程自动化 Robot Process Automation..... 1

 3.3 光学字符识别 Optical Character Recognition..... 1

 3.4 录音录像电子文件 Audio And Audio-Visual Electronic Document..... 2

 3.5 归档 Archiving..... 2

 3.6 实时音视频通信 Real-Time Communication; RTC..... 2

 3.7 实时音视频服务 RTC Service..... 2

 3.8 采样频率 Sampling Rate..... 2

 3.9 弱网 Weak Network..... 2

 3.10 比特率 Bit Rate..... 2

 3.11 包抖动 Packet Jitter..... 2

 3.12 丢包率 Packet Loss Rate..... 2

 3.13 数据生命周期 Data Lifecycle..... 2

4 略缩语..... 2

5 应用功能..... 3

 5.1 排队功能..... 3

 5.2 多人音视频功能..... 3

 5.3 单人音视频+AI 功能..... 3

 5.4 实时质检功能..... 3

6 基础能力..... 4

 6.1 音频能力..... 4

 6.1.1 编码协议..... 4

 6.1.2 音质处理..... 4

 6.1.3 采样频率..... 4

 6.2 视频能力..... 4

 6.2.1 编解码协议..... 4

 6.2.2 画质处理..... 4

 6.3 可用性能力..... 5

6.3.1 常规场景可用性保障.....5

6.3.2 弱网场景可用性保障.....5

6.4 智能化能力.....5

6.4.1 活体检测.....5

6.4.2 人脸检测能力.....5

6.4.3 人像比对能力.....5

6.4.4 公安校验.....6

6.4.5 智能选帧能力.....6

6.4.6 语音识别及合成能力.....6

7 数据存储与归档.....6

7.1 收集.....6

7.2 整理.....6

7.3 建立目录数据库.....6

7.4 归档与保存.....7

8 安全要求.....7

8.1 信息安全.....7

8.2 数据安全.....7

8.3 网络安全.....7

8.4 业务安全.....8

9 服务及技术创新要求.....8

10 实施保障.....8

10.1 组织保障.....8

10.2 管理制度.....8

前 言

本标准按照GB/T1.1—2020给出的规则起草。

本标准由长江证券股份有限公司提出。

本标准起草单位:长江证券股份有限公司。

本标准主要起草人:潘进、方军军、阮笛沛、余敏、陶慧明、瞿天卓、何晨曦。

引 言

实时智能音视频服务是证券公司在面向客户提供产品、服务过程中必不可少的能力之一。作为与客户实时交互的基础，智能音视频服务一旦出错，不仅会直接影响客户的服务体验，还会引发客户的隐私安全等问题。同时，智能音视频服务的质量、体验、技术先进性等也会直接影响相关业务办理的效率。为了进一步规范智能音视频服务，强化智能音视频服务安全与质量管理，编制本规范。

长江证券实时智能音视频技术规范

1 范围

本文件规定了实时智能音视频技术要求，具体包括应用能力、基础能力、数据处理、安全要求、服务创新，并确立了实施实时智能音视频服务的保障机制。

文件适用于实时智能音视频服务的设计、开发、实施、运维等相关工作。

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 5271.28-2001 信息技术 词汇 第28部分：人工智能 基本概念与专家系统
- GB/T 35273-2020 信息安全技术 个人信息安全规范
- DA/T 77-2019 纸质档案数字复制件光学字符识别（OCR）工作规范
- DA/T 78-2019 录音录像档案管理规范
- DA/T 63-2017 录音录像类电子档案元数据方案
- DA/T 79-2019 证券业务档案管理规范
- JR/T 0171-2020 个人金融信息保护技术规范
- JR/T 0223-2021 金融数据安全 数据生命周期安全规范
- JR/T 0071-2020 金融行业网络安全等级保护实施指引

3 定义和术语

3.1 人工智能 Artificial Intelligent

一门交叉学科，通常视为计算机科学的分支，研究表现出与人类智能（如推理和学习）相关的各种功能的模型和系统。

[来源：GB/T 5271.28-2001，28.01.01]

3.2 机器人流程自动化 Robot Process Automation

通过软件机器人基于一定规则的交互动作来模拟和替代人工对系统、软件应用、web等的操作，以自动执行响应的业务流程，加快流程执行效率。

3.3 光学字符识别 Optical Character Recognition

通过信息技术对图像文件中的字符形状进行识别、文字转换和文本输出、呈现的过程。

[来源：DA/T 77—2019，3.3]

3.4 录音录像电子文件 Audio And Audio-Visual Electronic Document

国家机构、社会组织或个人在履行其法定职责过程中,通过计算机、数字化转换等电子设备形成、传输和存储的数字音频和数字音视频文件。录音录像电子文件由内容、结构、背景组成。

[来源: DA/T 78—2019, 3.2]

3.5 归档 Archiving

将具有凭证、查考和保存价值并经系统整理的录音录像文件及其元数据管理权限向档案部门提交的过程。

[来源: DA/T 78—2019, 3.5]

3.6 实时音视频通信 Real-Time Communication; RTC

以服务端为中转节点,对终端用户的图像、视频、音频数据进行即时采集、渲染、处理、传输,从而实现音视频流数据在终端节点间完成通信的过程。

3.7 实时音视频服务 RTC Service

面向在线会议、直播营销、视频见证等应用,提供的实时音视频技术服务。

3.8 采样频率 Sampling Rate

单位时间内被采样处理的次数。

3.9 弱网 Weak Network

网络的复杂性、异构性、协议规范性等被破坏,所引起的网络异常、网络错误的情况。

3.10 比特率 Bit Rate

单位时间内传输或处理的比特的数量。

3.11 包抖动 Packet Jitter

描述包在网络中的传输延时变化,是包延时与平均传输延时的差值的平均值。

3.12 丢包率 Packet Loss Rate

测试中所丢失数据包数量占所发送数据组的比率。

3.13 数据生命周期 Data Lifecycle

数据从产生,经过数据采集、数据传输、数据存储、数据处理(包括计算、分析、可视化等)、数据交换,直至数据销毁等各种生存形态的演变过程。

[来源: GB/T 35274—2017, 3.2]

4 略缩语

下列缩略语表示适用于本文件:

AI: 人工智能(Artificial Intelligent)

AIGC: 生成式人工智能(Artificial Intelligent Generated Content)

AAC: 高级音频编码 (Advanced Audio Coding)
 ASR: 自动语音识别 (Automatic Speech Recognition)
 TTS: 文字转语音 (Text To Speech)
 OCR: 机器人流程自动化 (Robot Process Automation)
 AEC: 回声消除 (Acoustic Echo Cancel)
 ANS: 背景噪声抑制 (Automatic Noise Suppression)
 AGC: 自动增益控制 (Automatic Gain Control)
 ERLE: 回声返回衰减增益 (Echo Return Loss Enhancement)
 FPS: 平均帧率 (Frames Per Second)
 MOS: 平均意见得分 (Mean Opinion Score)

5 应用功能

实时音视频场景应用上, 应该包含排队功能、多人音视频通话功能、单人音视频录制功能、实时质量检测功能。

5.1 排队功能

排队功能应满足如下要求:

- a) 支持客户排队策略, 针对排队客户的属性进行排队策略、路由策略、分配策略的管理;
- b) 支持 VIP 插队策略, 按客户标签等级进行插队。

5.2 多人音视频功能

多人音视频通话功能, 应满足如下要求:

- a) 视频见证中, 支持客户通过移动 APP 终端与 PC 端坐席见证人员进行通话, 完成客户身份及意愿确认;
- b) 常规会议中, 支持不同终端进入同一个房间进行实时音视频通话;
- c) 支持两路或者多路音视频合成一路进行录制、录像上传功能;
- d) 支持集成 AI 厂商智能能力, 实现对双向视频见证的智能化。包括机器智能播报、业务办理语音确认、客户头像自动采集与识别、自动公安核查、视频自动录制、身份证质检以及客户信息核对功能, 提升视频见证规范与工作效率。

5.3 单人音视频+AI 功能

单人音视频录制功能, 应满足如下要求:

- a) 单人音视频录制中, 无需见证排队, 完成单人音视频的录制及上传;
- b) 支持客户端本地录制、服务端录制等录制模式;
- c) 支持音视频对接 AI 能力, 包含 ASR、TTS、人脸识别、活体检测等, 客户自助完成视频见证流程。

5.4 实时质检功能

实时质量检测功能, 为保障音视频通话质量, 应满足如下要求:

- a) 对实时音视频通话中视频质量指标进行实时监控, 包括视频分辨率、帧率、码率、压缩比、信

噪比等指标；

- b) 对实时音视频通话中音频质量指标进行实时监控，包括采样率、比特率、编码质量等指标；
- c) 对实时音视频通话中用户体验指标进行实时监控，包括卡顿次数、持续时间、MOS 值等指标。

6 基础能力

实时智能音视频的基础能力，即可支持面向音频、视频等数据的实时互动。其中核心能力包括音频能力、视频能力、可用性能力、智能化能力。

6.1 音频能力

实时智能音视频应保障语音通话的音质。

6.1.1 编码协议

实时音频支持主流的音频编码协议，应满足如下要求：

- a) 支持 Opus 编码协议；
- b) 支持 AAC 编码协议。

6.1.2 音质处理

实时音频应提供相应能力提高通话音质，音质处理应满足如下要求：

- a) 具备 AEC 能力，实现回声消除，减少用户在使用音视频过程中回声产生的负面影响，其中 AEC 的评价指标 ERLE 回声返回衰减增益推荐不低于 6db；
- b) 具备 AGC 能力，实现自动增益控制，缓解用户在使用音视频过程中因设备采集差异、说话人音量大小、距离远近等因素导致的音量的差异，防止爆音；
- c) 具备 ANS 能力，实现背景噪音抑制，减少用户背景噪音。

6.1.3 采样频率

实时音频应支持标准主流的采样策略，应满足如下要求：

- a) 支持 48kHz，全频段采样。

6.2 视频能力

6.2.1 编解码协议

实时视频应支持主流的视频编解码协议，应满足如下要求：

- a) 支持 H264 视频编码协议；
- b) 支持 H265 视频编码协议；
- c) 支持 VP8 视频编码协议；
- d) 支持 VP9 视频编码协议。

6.2.2 画质处理

实时视频需提供对应功能保证视频画质清晰，应满足如下要求：

- a) 应支持 1080p;
- b) 应具备算法策略, 平衡比特率和画质;
- c) 支持关键帧请求及超分算法。

6.3 可用性能力

保证实时音视频通信的基本可用性能力及可靠性, 可以由两个方面进行衡量, 一方面为指定分辨率情况下, 通话质量应达到对应指标, 另一方面为指定对应分辨率及质量指标的情况下, 可承受的网络弱化指标。

6.3.1 常规场景可用性保障

在常规场景下, 音视频的可用性应达到对应指标要求, 以在视频分辨率 1280*720、网络带宽充足情况为例, 应满足以下指标要求:

- a) 通话延迟应小于 500ms;
- b) FPS 平均帧率应大于 20;
- c) MOS 值应大于 4, 保持基本通话流畅。

6.3.2 弱网场景可用性保障

在网络带宽及网络稳定性有限的弱网情况下, 为保证音视频的稳定可用性, 需有对抗弱网相关能力, 以达到视频分辨率 320*240, 帧率 15, MOS 值大于 3 的实时音视频场景要求的情况下, 应可承受弱网的要求如下:

- a) 可承受最大丢包率应大于 30%;
- b) 可承受最大包抖动应大于 800ms;
- c) 满足要求最低带宽应小于 200Kbps;
- d) 通话延迟应小于 500ms。

6.4 智能化能力

6.4.1 活体检测

识别人脸照片、视频或者其他媒体资源是否来自活体, 即判断是否为真实人脸而非照片、视频等的技术能力。具备可靠的活体检测能力以提高识别的准确性和安全性。

6.4.2 人脸检测能力

识别图像或视频中人脸的技术能力, 要求如下:

- a) 能够识别多角度人脸姿态;
- b) 能够识别不同人种、年龄段、性别的人脸;
- c) 能够适应不同光照条件下的人脸检测, 如逆光、低光等;
- d) 能够同时处理图像或视频中多个人脸的检测;
- e) 能够采取相应的数据加密、隐私保护措施, 确保用户的人脸数据不被滥用或泄露;
- f) 能够应对一些常见的欺骗攻击, 如照片、视频等的作弊行为进行防范。

6.4.3 人像比对能力

根据用户提供的身份证照片，和现场视频中的人像进行比对，确定是否为同一个人。支持相似度阈值设置，可根据实际效果设置合适阈值。

6.4.4 公安校验

进行公安实名验证，人、证、卡、号一致性判断。

6.4.5 智能选帧能力

视频流中自动选取高质量满足要求的帧作为免冠照上传保存。需有一套算法准确性及其他指标的评估标准，要求算法在不同场景和数据集下具有良好的稳定性和一致性。

6.4.6 语音识别及合成能力

需满足智能问答、意愿确认等环节的需求。确定语音识别算法和模型的准确性评估标准，并发、支持语种、采样率、准确率等指标，需给出明确的评价。设定算法的响应时间要求，以满足实时性需求。在性能压测中明确压测的场景、需提供软硬件相关配置以及压测的结果。考虑语音数据的安全和隐私问题，并制定相应的保护措施。

7 数据存储与归档

对于公司履行职能活动中形成的、具有保存价值的音视频文件，应进行收集、整理、建立目录数据库、最后归档与保存。

7.1 收集

音视频文件收集阶段应符合要求：

- a) 音视频文件应客观、系统地反映主题内容，画面完整、端正，声音和影像清晰；
- b) 应以通用或开放格式收集、存储并归档音视频电子文件。音频电子文件归档格式为 WAV、MP3、AAC 等，视频电子文件归档格式为 MPG、MP4、FLV、AVI 等。

7.2 整理

应归档的音视频文件及记录载体的整理应符合以下基本要求：

- a) 以件为管理单位整理音视频电子文件，整理结果应能保持音视频电子文件之间的内在有机联系，通过规范命名、著录等建立音视频电子文件与目录数据之间的一一对应关系，便于利用和长期保存；
- b) 应按音视频文件记录的工作活动时间顺序排列音视频文件记录载体，按照规则为其编号并标示。

7.3 建立目录数据库

对应归档的音视频文件，基于电子档案管理系统进行著录，以对音视频文件记录的工作活动进行描述，建立目录数据库。

应包含如下信息：

- a) 文件描述信息：包括文件名，文件大小，文件日期等；
- b) 音视频元数据信息：包括音视频摄录日期、时间长度、编码标准、比特率、视频分辨率、高宽比、帧率、音频采样率、声道等；
- c) 工作活动信息：包括工作活动名称、起始日期、地点及活动描述等；

- d) 保管期限:音视频文件保管期限的代码,永久—Y,定期 30 年—D30;
- e) 密级:音视频文件的保密等级,分为商业秘密一级、商业秘密二级和商业秘密三级;
- f) 原始载体编号:记录或存储音视频文件的原始载体编号,多个原始载体编号之间用“,”隔开。

7.4 归档与保存

归档与保存要求但不限于:

- a) 自音视频文件形成起 3 个月内向档案部门提交归档,最迟不能超过音视频文件形成后的次年 6 月;
- b) 可采用在线和离线两种归档方式。基于电子档案管理系统整理、索引的音视频电子文件,以在线方式归档;记录或存储音视频文件、目录数据的原始载体以离线方式归档;
- c) 分为 30 年和永久两种保管期限。

8 安全要求

8.1 信息安全

长江证券智能音视频服务在信息安全的要求如下:

- a) 应在客户账户信息、鉴别信息、金融交易信息、个人身份信息、财产信息、借贷信息和个人信息保护等方面符合 JR/T 0171—2020《个人金融信息保护技术规范》和 GB/T 35273—2020《信息安全技术个人信息安全规范》;
- b) 音视频个人信息采集上,应依据最小化原则,仅收集为用户提供产品或服务有直接关联的个人信息,不应收集与该产品或服务无关的个人信息。收集信息的频率不应超出产品或服务所必需的频率;
- c) 应明确说明个人信息的查询、更正、删除、账户注销、撤回的操作方法,并提供电子邮件、电话、传真、在线客服等投诉渠道;
- d) 应对通过音视频界面展示的个人敏感信息采取字段屏蔽等处理措施,降低个人敏感信息在展示环节的泄露风险。

8.2 数据安全

为防范和抵御智能音视频的数据安全风险,在提供智能音视频服务的过程中,遵循以下数据安全基本要求:

- a) 应确保音视频数据全生命周期各环节数据活动的合法性和正当性;
- b) 应制定音视频数据安全防护策略,明确音视频数据生命周期各环节的安全防护目标和要求;
- c) 应采取与音视频数据安全级别相匹配的安全管控机制和技术措施,确保数据在全生命周期各环节的保密性、完整性和可用性,避免数据在全生命周期内被未经授权访问、破坏、篡改、泄漏或丢失等;
- d) 音视频数据的安全控制策略和安全防护措施不应是一次性和静态的,应可基于业务需求、安全环境属性、系统用户行为等因素实施实时和动态调整;
- e) 数据采集、传输、存储、访问、删除应满足 JR/T 0223-2021《金融数据安全 数据生命周期安全规范》中的安全要求。

8.3 网络安全

网络安全方面,应满足如下要求:

- a) 网络安全等级保护应参照 JR/T 0071—2020《金融行业网络安全等级保护实施指引》;

- b) 应在关键网络节点处检测、防止或限制从内部发起的网络攻击行为；
- c) 应采取技术措施对网络行为进行分析，实现对网络攻击特别是新型网络攻击行为的分析。当检测到攻击行为时，应记录攻击源 IP、攻击类型、攻击目标，攻击时间，在发生严重入侵事件时应及时报警；
- d) 应建立诱捕、欺骗攻击者的安全防护手段，对攻击者的行为进行捕获和分析；
- e) 网络安全监控，应划分出特定的管理区域，对分布在网络中的安全设备或安全组件进行管控，对网络链路、安全设备、网络设备和服务器等运行状况进行集中监测，对安全策略、恶意代码、补丁升级等安全相关事项进行集中管理。

8.4 业务安全

业务安全方面，应满足如下要求：

- a) 应确保会话标识随机并且唯一，可使用中间件系统提供的会话标识生成机制；
- b) 应在会话过程中维持认证状态，防止用户通过直接输入登录后的地址访问页面；
- c) 应针对涉及资金变动的交易和客户签约信息变更的交易实施防抵赖保护；
- d) 智能音视频服务应开启、记录和保存安全审计日志，覆盖到每个用户的安全审计功能，对重要的用户行为和重要安全事件进行审计。

9 服务及技术创新要求

本项要求包括但不限于：

- a) 应针对单客户提供音视频服务，包括业务办理、VIP 服务等，结合先进音视频传输与处理技术，实现音视频服务低时延、高稳定、高质量；
- b) 应针对多客户同时提供音视频服务，包括交流会、宣讲会等，更加方便、快捷地开展业务；
- c) 结合音视频和语音识别及合成技术，以虚拟数字人的形式面向客户提供服务，更加智能化的提供答疑、流程指引等功能，并结合 AIGC 技术，丰富音视频服务内容，给用户沉浸式体验；
- d) 应通过音视频技术，结合身份鉴别、声纹识别、人脸识别等，实现客户高精度无感识别，提升认证效率与安全性；
- e) 应结合图像识别、文字识别等技术，提高音视频服务过程安全性，实现自动预警等功能。

10 实施保障

10.1 组织保障

为确保智能音视频服务创新和研发工作的顺利开展，研发组织的设立要求包括但不限于：

- a) 规划目标清晰，且符合公司战略要求；
- b) 具备专门的产品团队，对智能音视频的应用进行研究和规划，并设计产品方案；
- c) 具备专门的资源调度和仲裁组织，负责审议资源规划，协调和调度可用资源；
- d) 具备专门的技术管理部门，负责审议智能音视频相关的解决方案、技术方案；
- e) 具备专门的开发团队，负责智能音视频相关的架构设计、技术选型和开发工作；
- f) 具备专门的测试团队，统筹管理测试工作，确保测试工作有序开展。

10.2 管理制度

本项要求包括但不限于：

- a) 智能音视频服务管理应遵守长江证券内部控制、合规管理、风险管理、信息安全等各项规章制度；
- b) 应制定智能音视频产品研发类制度，遵循长江证券研发规范，包括项目开发规范、设计规范、评审机制等；
- c) 应制定智能音视频产品测试类制度，遵循长江证券测试规范，包括测试过程规范、测试质量管理规范等；
- d) 应制定智能音视频生产运营类制度，遵循长江证券生产运营规范，包括配置规范、变更规范等；
- e) 应制定完备的智能音视频管理制度，包括服务章程、服务协议、内部操作流程等。