团体标准《信息技术 智能媒体编码 第10部分：

实时语音》（征求意见稿）编制说明

## 工作简况

### 任务来源

根据中关村视听产业技术创新联盟2024年标准制修订计划的安排 下达了团体标准《信息技术 智能媒体编码 第10部分：实时语音》的制定任务，标准计划号为：2024010801，牵头起草单位包括清华大学、腾讯科技（深圳）有限公司、中国电子技术标准化研究院、华为技术有限公司等。

### 主要工作过程

《信息技术 智能媒体编码 第10部分：实时语音》（以下简称AVS3P10实时语音）的制定工作，起始于2023年6月举行的数字音视频编解码技术标准工作组（简称AVS工作组）第85次会议，根据AVS音频组和需求组联席会议确定了AVS语音编码的技术需求V1.0(N3577)，发出“AVS语音编码技术方案征集书V1.0”(N3596)，启动面向实时语音通信场景的低码率高质量语音编码项目技术的征集工作。2023年8月根据新的技术需求提案，发布了更新的AVS3P10实时语音需求V1.1(N3630)和相应的技术方案征集书V1.1(N352)。2023年12月基于已采纳的技术提案，形成了AVS3P10 实时语音WD1.0。2024年1月8日，经新一代人工智能产业技术创新战略联盟和中关村视听产业技术创新联盟标准工作组联合审议决定立项制定AVS3P10实时语音团体标准。2024年 3月基于新采纳的技术提案，形成了AVS3P10实时语音CD1.0(N3758)。2024年5月，对照参考软件，对CD1.0进行了勘误，形成了CD2.0(N3814)。此后，通过文本校对、编解码器交叉验证、参考软件与标准文本的一致性检验等工作，于2024年6月，形成AVS3P10实时语音的FCD(N3842)，并形成此征求意见稿。截止2024年6月15日，历经AVS工作组正式和临时加会议10次，共收到AVS3P10实时语音相关提案27项，被采纳的技术提案3项，收到第三方主观测试报告和交叉验证报告各2份。按照AVS音频组工作规范，AVS音频组对所有技术提案和信息提案均进行了认真审议和充分论证，经过从技术需求、技术征集书、技术提案、测试方案、测试报告、参考软件验证、工作组草案WD、委员会文档CD、最终文档FCD等完整的标准研制过程，形成了AVS3P10实时语音标准征求意见稿。

### 主要参加单位和制定人员及所做工作

AVS3P10实时语音标准的主要参加单位是：清华大学、腾讯科技（深圳）有限公司、中国电子技术标准化研究院、华为技术有限公司、北京理工大学、北京全景声信息科技有限公司、中关村视听产业技术创新联盟、清华大学天津电子信息研究院。

AVS3P10实时语音标准的主要起草人是：窦维蓓、肖玮、张亚军、商世东、李婧欣、朱博成、王喆、王晶、许舒敏、潘兴德、张伟民、高文、黄铁军。

主办单位清华大学的窦维蓓是数字音视频编解码技术标准工作组（AVS工作组）音频组组长，作为本标准制定工作的牵头人，在主办单位清华大学和清华大学天津电子信息研究院的大力支持下，依托AVS工作组提供的管理平台和协调服务，在AVS音频组召集国内实时语音领域有经验的高校、企业、事业单位参加本标准的制定，并负责组织协调整个标准制定工作，包括：申请标准立项，制定工作计划和工作规范，组织技术需求研讨、技术征集和技术提案审议，协调技术方案测试验证，提出标准文本的整体架构，主笔部分文本内容编写，负责准备标准文本送审和报批等工作。

主要起草人腾讯科技（深圳）有限公司的肖玮、张亚军和商世东，协助牵头人负责召集需求研讨、提案征集和审议，承担技术方案的设计实现、主笔标准文本和参考软件维护，组织标准文档编辑、标准文本征求意见和意见答复，承担文本校对和格式审查等工作。主要起草人中国电子技术标准化研究院的李婧欣和朱博成，华为技术有限公司的王喆，和北京理工大学的王晶主要负责技术方案的测试验证。主要起草人北京全景声信息科技有限公司的许舒敏和潘兴德，组织对技术内容和参考软件的交叉验证，承担文本校对和格式审查等工作。主要参加单位中关村视听产业技术创新联盟，通过数字音视频编解码技术标准工作组（AVS工作组）为本标准制定提供了必备的管理交流平台和协调服务工作。

## 标准编制原则和确定主要内容的论据及解决的主要问题

### 标准制定原则

在本标准制定过程中，标准制定工作组充分考虑到标准自身的科学性、合理性和易操作性，同时也为了使标准使用者易于理解标准内容，将满足对标准技术内容完整准确表达的前提下进行标准内容制定，力图达到语言表达形式尽可能简单、易懂，使相应专业人员便于理解和使用。

在编写内容与格式上按GB/T 1.1-2020 《标准化工作导则 第1部分：标准文件的结构和起草规则》要求进行编写。

### 标准制定主要内容

本标准，面向下一代实时语音通信应用场景，适用于在线会议、协同办公、社交、互动娱乐等领域。

本标准定义了实时语音编码标准相关的术语和定义。特别地，包含了与深度神经网络相关的术语和定义。

本标准定义的实时语音编码是一种基于子带编码的多模式编码方案。按声道数划分，支持单声道和立体声两种编码模式；按采样频率划分，支持宽带（fs=16000Hz）和超宽带（fs=32000Hz）两种模式。本标准定义的实时语音编码和解码框架见图1和图2。



图1 AVS3P10实时语音编码框架



图2 AVS3P10实时语音解码框架

本标准定义了实时语音解码工具，见第7章。内容包含：概述、原始位流数据、宽带特征向量信息和宽带特征向量、解码神经网络、MCR立体声解码和带宽扩展解码。

附录A包含ASTF语法及语义，附录B包含解码神经网络信息，附录C包含AVS3P10实时语音编码参考编码器，附录D包含丢包隐藏。

## 主要试验[或验证]情况分析

AVS3P10实时语音编码标准中采纳的所有技术均依据AVS3P10实时语音测试方案进行了第三方测试验证。测试依据包含ITU-T P.800 DCR（Rec. P. 800: Methods for subjective determination of transmission quality. (1996)[J]. Geneva, Switzerland: ITU）。

AVS3P10实时语音测试结果（见编制说明附录）表明，本标准在技术和性能上达到先进水平，符合标准技术需求，应用场景广泛，满足标准发布条件。

## 知识产权情况说明

AVS3P10实时语音标准制定参与单位声明专利4项，涉及的单位包括：腾讯科技（深圳）有限公司、清华大学、华为技术有限公司。相关专利权人均同意其声明之专利加入AVS专利池。

标准中涉及的专利信息如下:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 序号 | 专利申请号/专利号  | 专利名称 | 专利申请人/专利权人 | 标准条款涉及专利的（章、条编号） |
| 1 | 202210681816.X | 音频编码及解码方法、装置、设备、介质及程序产品 | 腾讯科技（深圳）有限公司 | 第7章附录A、B、C |
| 2 | 202210681365.X | 音频处理方法、装置、设备、存储介质及计算机程序产品 | 腾讯科技（深圳）有限公司 | 第7章附录A、B、C |
| 3 | 202210676984.X | 音频解码、编码方法、装置、电子设备及存储介质 | 腾讯科技（深圳）有限公司 | 第7章附录A、B、C |
| 4 | ZL 200810106460.7 | 立体声信号编解码方法、装置及编解码系统 | 华为技术有限公司、清华大学 | 7.5附录C.5-6 |

## 产业化情况、推广应用论证和预期达到的经济效果

AVS3P10实时语音标准具有自主知识产权、技术水平达到或超过行业标杆，在在线会议、社交等领域得到了应用落地。例如，腾讯科技（深圳）有限公司已经将与AVS3P10实时语音语音相关的技术应用于腾讯会议和QQ等多个音视频通话产品。此外，腾讯科技（深圳）有限公司，采用本标准技术方案，于2024年6月13-15日在绍兴举行AVS工作组第八十九次会议期间，基于腾讯会议应用软件，现场展示了AVS3P10实时语音标准实时通话效果，展示特性包括：同等质量下更低位率、同等位率下更高质量、强噪声干扰下的高清通话效果等。良好的体验效果标志着AVS3P10实时语音标准在实时语音通信的应用产品中成功落地。

本标准将与实时语音通信相关行业发展相配合，推动在线会议、协同办公、社交、互动娱乐等领域，为用户带来更好的性能和体验。同时，促进人工智能技术融入和应用到实时语音通信领域中，带动相关产业的经济效益和社会效益。

## 采用国际标准和国外先进标准情况

传统的语音编码器，包括3GPP EVS、IETF OPUS等标准语音编码器，在16-20kbps左右位率时，能够恢复出高质量宽带语音；在30-35kbps，可以恢复出高质量超宽带甚至全带语音。然而，当位率进一步降低（如：降到10kbps以下时），传统语音编码器恢复的质量下降明显，影响用户体验。AVS3P10实时语音编码标准，成功解决了10kbps以下通话质量不足的问题，见附录AVS3P10实时语音测试结果。

与此同时，IETF Applications and Real-Time Area工作组于2023年成立一个新的AI codec标准工作组Machine Learning for Audio Coding WG（IETF MLCodec）。其目标是围绕IETF OPUS生态，通过引入深度神经网络技术，提升OPUS的质量。该工作预计在2024年下半年完成标准化工作。2024年第一季度，该标准已经完成第一阶段工作；但技术指标上，宽带语音降到10kbps以下时，主观质量与本标准有明显差距，见附录AVS3P10实时语音测试结果。

## 与现行相关法律、法规、规章及相关标准的协调性

本文件符合我国有关的现行法律、法规。

## 重大分歧意见的处理经过和依据

无。

## 标准性质的建议

建议本标准为推荐性标准。

## 贯彻标准的要求和措施建议

建议按照专家意见修改送审稿尽快完成编制以及标准的最终发布，可以更好引导产业发展，规范市场，正确引导消费者。

## 替代或废止现行相关标准的建议

无。

## 其它应予说明的事项

无。

《信息技术 智能媒体编码 第10部分：实时语音》

编制工作组

2024年6月18日

附录：AVS3P10实时语音测试结果

AVS3P10实时语音进行了第三方测试验证，回放工具为耳机。测试内容主要是评估基本音质，即相对参考信号声音品质上的变化。AVS3P10实时语音部分性能测试结果如下：



图3 单声道主观测试结果



图4 立体声主观测试结果

主观测试结果表明，在单声道编码场景，AVS3P10实时语音在宽带和超宽带等多个主要测试场景均达到了4.0以上MOS分，体现出明显优势，最低位率可达5.9kbps。AVS3P10实时语音采纳了深度神经网络技术，自带的丢包损伤能力，有效提升了编码器在网络不佳时的质量。在立体声编码场景，位率低至7.6kbps时也可实现高质量双声道立体声编码效果，主观质量均达到4.0以上MOS分。在低于17kbps的测试位率条件下，AVS3P10实时语音实现了双声道立体声主观质量均超过对比系统。



图5 与IETF标准测试结果

此外，基于ITU-T P.863客观语音质量评价标准，对比了AVS3P10实时和IETF MLCodec目前输出的版本（包括Opus、Opus\_Lace和Opus\_NLace）的质量。从结果看，AVS3P10项目在6-10kbps，质量有显著优势。这充分反映了AVS3P10实时语音标准的技术领先性。