

多设备协同场景下音视频同步技术研究

赵洪良¹, 张平贺^{2*}

1. 中国电子技术标准化研究院, 北京 100007

2. 工业和信息化部教育与考试中心, 北京 100400

DOI: 10.61369/TACS.2026050016

摘 要 : 随着多设备持有率提升与多媒体应用普及, 音视频交互已广泛应用于远程会议、多机位录制、在线教育等场景, 音视频同步精度成为影响用户体验的核心指标。基于此, 本文立足多设备协同场景特性, 梳理音视频同步技术的核心原理与现状, 重点探讨音视频同步技术优化策略, 以期推动该技术向高效化、智能化方向演进。

关 键 词 : 多设备; 数字化; 音视频同步; 交互体验; 稳定

Research on Audio-Video Synchronization Technology in Multi-Device Collaboration Scenarios

Zhao Hongliang¹, Zhang Pinghe^{2*}

China Electronics Standardization Institute, Beijing 100007

Education and Examination Center of the Ministry of Industry and Information Technology, Beijing 100400

Abstract : With the increase in multi-device ownership and the popularization of multimedia applications, audio-video interaction has been widely used in scenarios such as remote conferences, multi-camera recording, and online education. Audio-video synchronization accuracy has become a core indicator affecting user experience. Based on this, this paper focuses on the characteristics of multi-device collaboration scenarios, sorts out the core principles and current status of audio-video synchronization technology, and focuses on exploring optimization strategies for audio-video synchronization technology. It aims to promote the evolution of this technology towards a more efficient and intelligent direction.

Keywords : multi-device; digitalization; audio-video synchronization; interactive experience; stability

引言

在数字化与智能化快速发展的背景下, 多设备协同已成为多媒体交互的主流形态。音视频同步技术作为多设备协同场景的核心支撑, 广泛应用于远程办公、多机位直播、智慧教育、沉浸式娱乐等多个领域, 直接决定用户交互体验的优劣。随着5G、物联网技术的普及, 多设备协同场景日益复杂, 现有音视频同步技术多针对单一设备或简单场景设计, 难以适配多设备协同的复杂需求, 同步精度与稳定性仍有较大提升空间^[1]。因此, 在多设备协同场景下, 优化音视频同步技术, 可以解决多设备协同中的音视频失步难题, 推动多媒体交互技术向更高效、更稳定、更智能的方向发展。

一、多设备协同场景下音视频同步技术的核心原理

(一) 时间戳同步原理

在多设备协同场景中, 不同设备在音视频的录制、编解码和数据传输等环节存在显著差异。若缺乏统一的时间标准, 音视频数据的产生时间和播放时间将无法对齐, 进而出现声音和画面不同步的问题。生成时间戳需要遵循采集端基准原则^[2]。简单来说, 在音视频数据刚采集完成的那一刻, 由采集设备为其分配唯一的时间戳标记, 精准对应数据实际生成的时间, 避免因采集延迟、编解码耗时这些问题造成时间戳偏移。

同时, 时间戳的格式需采用统一标准, 既要保证精度, 又要

兼顾兼容性, 在满足毫秒级的同步要求的同时, 还要能适配不同设备、不同编解码标准, 让多设备之间的时间戳可以互相识别、互相比较。在音视频数据传输的过程中, 时间戳需与音视频帧数据绑定传送, 确保数据到达接收端后, 接收设备能够快速解析时间戳, 明确数据的时序关系^[3]。

(二) 时钟同步原理

在多设备协同场景中, 音视频同步的本质是多设备间音视频数据时序的对齐, 而时钟偏差是导致时序错位的核心诱因, 需建立统一、精准的时间基准, 促使多设备在音视频采集、传输、播放全流程的时间一致性。在多设备协同场景中, 集中式时钟同步模式以一个主设备作为时钟基准, 其他从设备通过与主设备的时

钟交互，校准自身时钟，实现时钟同步。在此过程中，主设备需要搭载高精度时钟模块，生成稳定可靠的参考时钟^[4]。从设备需定期向主设备发送时钟请求，获取参考时钟信息，再调整自己的时钟频率和相位，以此实现设备间的时钟同步。而分布式时钟同步模式，是依靠各个设备之间相互交互时钟信息，搭建分布式的同步网络。在这种模式下，每台设备既是时钟的发送方，也是时钟的接收方，通过多轮交互与计算，获取整个网络统一的参考时钟，再以此校准自身时钟。

（三）缓冲与延迟补偿原理

在多设备协同场景中，音视频数据在采集、编解码、传输、播放等环节会产生不同程度的延迟，且延迟具有随机性、动态性特征，而缓冲机制的核心作用是缓存音视频数据，调节数据传输与播放的速率差，避免因传输延迟波动导致的播放卡顿或失步^[5]。一方面，根据多设备协同场景的延迟特征，设计动态缓冲容量调整机制，实时监测数据传输延迟的变化，动态调整缓冲容量，在保证播放流畅性的同时，降低播放延迟，实现实时性与流畅性的平衡；另一方面，精准监测各类延迟的实时值，通过延迟检测算法，计算多设备间音视频数据的延迟差值，进而通过调整播放时序、优化数据传输速率等方式，对延迟差异进行补偿。

二、多设备协同场景下音视频同步技术优化困境

（一）核心同步技术瓶颈凸显，难以支撑高精度协同需求

当前各类核心技术的短板相互叠加，形成了难以逾越的技术壁垒，无法充分适配多设备异构、场景复杂的实际应用场景。第一，多设备协同场景中，不同设备的硬件架构、操作系统、编解码标准存在本质差异，导致各设备的时间生成基准存在固有偏差，而时间戳校准技术存在明显短板，难以实现多设备时间基准的绝对统一，导致同步精度下降^[6]。

第二，从时钟同步技术来看，集中式时钟同步模式受主设备性能与负载限制，难以适配大型分布式多设备协同场景，且主设备的单点故障会直接导致整个时钟同步体系崩溃，缺乏足够的可靠性与扩展性。同时分布式时钟同步模式虽具备较强的扩展性与容错能力，但现有同步算法难以有效抵消这些干扰，导致同步精度难以满足高精度场景的应用需求。第三，现有缓冲机制多采用固定容量阈值设计，缺乏对延迟波动的实时感知与动态调整能力，从而出现播放延迟、卡顿失步的问题。

（二）技术融合程度不足，未能形成协同赋能效应

当前，5G、人工智能、物联网、边缘计算等新兴技术的快速发展为同步技术优化提供了新的支撑，但同步技术与5G、物联网技术的融合存在表层化问题，未能充分发挥两类技术的核心优势。在更新和优化过程中，同步技术未充分结合5G网络的低延迟、高带宽、高可靠的特性，仅实现了数据传输层面的简单适配，难以提升多设备协同音视频同步精度^[7]；此外，将人工智能技术简单应用于延迟预测与参数调整，未能构建起全流程智能化同步优化体系，机器学习算法的训练数据覆盖面有限，难以精准捕捉多设备协同场景下延迟变化、设备状态波动的复杂规律。

（三）跨场景适配能力薄弱，应用拓展受到制约

随着多设备协同应用在各个领域越来越普及，不同场景对音视频同步的精度、实时性还有稳定性需求存在显著差异。当前同步技术大多采用通用化的设计思路，没有根据不同场景的核心需求做针对性优化^[8]。这样一来，导致同步技术在部分场景里难以充分发挥出它应有的效果。对于要求高精度、低延迟的专业场景，通用同步技术的精度与响应速度难以满足需求。另外，不同场景下的同步技术优化标准不统一，缺乏针对性的优化规范，导致同步技术在跨场景应用时，需要进行大量的适配调整，难以适配跨场景高精度需求。

三、多设备协同场景下音视频同步技术优化策略

（一）构建核心同步体系，夯实高精度同步基础

核心同步技术是多设备协同场景下音视频同步系统的核心支撑，需遵循精准化、动态化、标准化原则，结合多设备协同场景的异构性、动态性特征，构建高精度、高稳定、高适配的核心同步体系，为多设备协同音视频交互提供坚实的技术支撑^[9]。

其一，建立统一的时间戳生成与校准标准。一方面优化时间戳分配算法，将时间戳生成与音视频采集瞬间精准绑定，最大限度降低采集、编解码等环节耗时对时间戳精度的影响；另一方面引入动态校准算法，融合网络时间协议与精确时间协议的技术优势，实现对多设备时钟偏差的实时检测与动态修正，提升校准算法的自适应能力。

其二，构建集中式与分布式协同互补机制。从集中式模式来看，需优化主设备时钟性能与从设备同步响应机制，提升主设备时钟的稳定性与精度，优化从设备的同步响应速度，减少同步延迟，同时增设主设备冗余备份机制，提升同步系统的可靠性与抗故障能力；从分布式模式来看，需优化设备间时钟交互算法，引入抗干扰设计，有效抵消设备交互延迟、网络抖动等外部因素的影响，提升同步精度^[10]。

其三，构建动态缓冲容量调整机制。实时监测网络状态、设备负载等动态因素的变化，精准预测延迟波动趋势，灵活调整缓冲队列容量，实现音视频交互的实时性与流畅性协同提升。

（二）推动技术深度协同，提升同步系统的自适应能力

5G、人工智能、物联网、边缘计算、云计算等新一代信息技术的快速发展与应用，为音视频同步技术的优化提供了新路径，将新兴技术与同步技术进行融合，实现技术优势互补、效能叠加，提升同步系统的自适应能力、优化效率与扩展能力^[11]。首先，搭建多维度、全方位数据采集体系，实时收集各个设备的运行情况、负载情况、网络状态等核心数据。然后将收集到的实时数据和同步参数优化深度结合，根据这些信息动态调整同步策略，提高同步系统的自我适应能力。并根据设备与网络的动态变化，自动优化同步参数，适配多设备协同场景的复杂变化，充分发挥物联网技术的感知赋能作用，提升同步系统的稳定性与适配性。

其次，借助机器学习、深度学习等算法，对历史延迟数据、

设备运行数据、网络状态数据进行深度分析挖掘，精准捕捉多设备协同场景下延迟变化、设备状态波动的复杂规律，构建高精度延迟预测模型，实现延迟变化的精准预判，提前规避延迟差异带来的时序错位^[12]。同时利用人工智能技术优化缓冲队列管理与时间戳校准算法，提升同步精度与系统运行效率，实现同步技术从数据采集、延迟预测到参数优化、时序对齐的全流程智能化升级。最后，建立同步技术与云计算协同机制，优化边缘计算与云计算的协同架构，实现边缘节点的同步计算任务与云端的集中管理无缝衔接，将部分同步计算任务下沉至边缘节点，减少云端传输延迟，实现对本地多设备同步的快速校准，提升同步响应速度。

（三）强化场景化适配，拓展技术应用边界

随着智能终端设备的普及，多设备协同被广泛应用于各个场景，音视频同步技术性能直接决定用户交互感受与系统服务质量，因此需立足不同多设备协同场景的差异化需求，开展个性化、针对性的优化设计，构建场景化同步优化方案，提升同步技术的跨场景适配能力。针对多设备协同场景的核心需求，明确同步精度、实时性、稳定性的差异化指标，制定针对性的优化标

准，确保同步技术的优化与场景需求精准匹配。比如对于对同步精度、实时性要求较高的场景，需重点优化同步精度与响应速度，强化时间戳校准、时钟同步与延迟补偿的精度，采用高精度同步协议与算法，实现微秒级同步，满足场景的高精度需求。

此外，总结不同场景的同步优化经验，提炼共性优化模块与核心算法，构建场景化同步优化模板，当同步技术向新场景迁移时，可基于模板进行个性化调整，减少重复开发，提升迁移效率。同时收集不同场景的优化数据、策略与效果，为后续场景化优化与技术迁移提供参考，推动同步技术的场景化适配持续深化。

四、结语

综上所述，音视频同步是多设备协同场景中最核心、最典型的技术指标之一，是否具备良好的同步性将会对多设备协同质量产生决定性的影响。通过构建核心同步体系、推动技术深度协同、强化场景化适配等策略，可以推动同步技术向高精度、智能化方向升级。

参考文献

[1] 章卫峰. 音视频同步技术在广播电视节目制作中的应用 [J]. 电声技术, 2025, 49 (02): 131-133.
[2] 张桁. 基于物理模型的火灾音视频同步校正与火源定位研究 [J]. 消防界 (电子版), 2024, 10 (16): 40-42.
[3] 王媛媛. 基于 PTP 的音视频同步技术在电视新闻多机位编辑中的应用 [J]. 电视技术, 2024, 48 (05): 48-50.
[4] 曾文杰. 面向无线投屏场景的音视频传输策略研究与实现 [D]. 武汉理工大学, 2024.
[5] 赵标. 多点视频会议音视频同步技术优化 [J]. 中国宽带, 2024, 20 (04): 61-63.
[6] 刘帅凤, 赵少雄, 周文硕. IP 视频会议系统中音视频同步的研究 [J]. 无线互联科技, 2024, 21 (07): 113-115.
[7] 陈航. 面向音视频语音增强与识别的建模方法研究 [D]. 中国科学技术大学, 2024.
[8] 杜宸翌, 李博, 画芊昊. 基于 MPC 的音视频同步码率自适应测试 [J]. 计算机测量与控制, 2024, 32 (04): 54-60.
[9] 洪学敏, 张海翔. 基于 LSTM-CBAM 的音视频同步人脸视频生成 [J]. 智能计算机与应用, 2023, 13 (05): 151-155.
[10] 方芳. 8K 超高清电视节目音视频时码同步的探索 [J]. 现代电视技术, 2022, (10): 48-51.
[11] 莫瑶. 车载娱乐系统音视频同步技术研究与应用 [D]. 电子科技大学, 2022.
[12] 袁金保, 黄化吉, 杨雪松. 一种直通音频与视频重同步的设计与应用 [J]. 电视技术, 2021, 45 (06): 112-114.