Holodeck: A Research Framework for Distributed Multimedia Concert Performances

Andrea F. Genovese[1], Zack Nguyen[2], Marta Gospodarek[1], Robert Pahle[2], Corinne Brenner[3], Agnieszka Roginska[1]

New York University, [1]*Music and Audio Research Lab*, [2]*Research Technology*, [3]*CREATE Lab*, New York, US.
genovese@nyu.edu, gospodarek@nyu.edu, zack.nguyen@nyu.edu, cjb399@nyu.edu, pahle@nyu.edu, roginska@nyu.edu

***Abstract*—This paper presents the Holodeck project, a multi-room research framework designed to support distributed multimedia concert performances. Developed through an inter-lab collaboration at New York University, the Holodeck platform leverages the Corelink engine, a flexible and unified data routing system that facilitates immersive, interactive experiences across diverse networked environments. The framework enables real-time streaming and synchronization of various data types, including audio, video, and motion-capture data, thus supporting the design and implementation of augmented Network Music Performances (NMPs). Discussion around case studies of large-scale distributed concerts illustrates the system's capabilities and potential applications. In a pilot research exploration, the platform has been used to gather preliminary data on the quality of experience from both audiences and musicians. This research aims to enhance telepresence and realism in remote collaborations, contributing to the development of new methodologies and artistic practices in multimedia performances. The paper also discusses the technical challenges and solutions associated with implementing such a flexible and adaptive system and future work regarding the platform and its usage in multimedia immersive NMPs.**

***Index Terms*—distributed music, interactive displays, multimedia, augmented concerts, XR**

I. Introduction

Network Music Performances (NMPs) occur when two or more music performers collaborate from separate places by means of telecommunication technology. The past two decades saw an increasing popularity of NMP practices, partly thanks to advancements in high-speed academic networks with reduced latency properties. Further interest in NMP surged during the COVID-19 pandemic due to travel restrictions [1], amplifying research efforts among a growing field intersecting network technology and music systems [2], [3]. As of today, several specialized software platforms that can handle both video and audio streams are available to the general public [4], facilitating remote collaborative interactions among artists, such as the enactment of distributed concerts [5].

These existing tools mostly focus on the efficient handling of multichannel audio streaming and broadcasting across multiple nodes and occasionally allow for companion streams consisting of two-dimensional video. However, these tools are also usually tailored for a very specific interaction paradigm and fixed layout and are not flexible towards alternative kinds of data that may be part of future kinds of augmented NMPs. There is in fact growing interest in the incorporation of immersive technology with NMP displays, for example by introducing eXtended Reality (XR) headset displays, spatial audio rendering, shared navigable virtual environments, and acoustic auralization, for improving the experience realism and telepresence [6]. There is thus a strong interest in the development of a flexible unitary environment that can allow diverse data types (e.g. motion-capture data) to be packaged and synced with the audio streams in a customizable way. This should take into account diverse needs and asymmetric properties of every connected node, optimize the data exchange lines to each node's capabilities, and process the data on the fly to best adapt and fit to the user's environment.

In this paper, the authors present the current progress on the *Holodeck* collaborative infrastructure, a result of an inter-lab collaboration across New York University. Holodeck is a multi-room research platform for developing multimedia, interactive, immersive experiences, tackling applications such as augmented distributed live music performances in small and large scale. This platform, and its core data exchange engine *Corelink*[1], have the potential to establish a reconfigurable flexible research framework that includes (but is not limited to) the design and study of NMP performances.

Firstly, the paper reviews tools and general limitations in the creation of multimedia immersive NMP environments, and provides an overview of the Holodeck infrastructure and the Corelink streaming routing tools. The discussion is followed by a series of case studies illustrating the design and implementation of example large-scale distributed multimedia live concerts that occurred under as series of events called the *Holodeck Concert Series*, including an exploratory study on the Quality of Experience, and an illustration of the compatibility of the system with external development frameworks.

The paper can be of particular interest to the growing community on the "Musical Metaverse" [7], [8], in particular to the branch of study linked to XR multimedia platforms, due to the number of interactive media elements that can interconnect flexibly at each node. In regards to NMPs, the utility of this framework is to drive the proliferation of

The work described in this paper has been partially funded by the NSF MRI Award #1626098

[1]Corelink is an open-source project available to the public, the authors encourage participation to the code repository.
Corelink repository link: https://dev.hsrn.nyu.edu/corelink

research-oriented concert events and promote work to help with the validation of novel evaluation methodologies and artistic techniques, discussing current challenges and tracing guidelines for future improvements.

II. Background on Multimedia NMPs

In the setup of an NMP network, there is usually a trade-off between latency amount, which is never fully eliminated, and audio quality. The NMP interaction design process is heavily influenced by geographic distance, network infrastructure, and available bandwidth for data transmission. The primary goal in NMP design is usually to keep one-way latencies within "natural" levels, ideally around 10-20ms. Above the range of 20-30ms of latency, musical performance is increasingly impaired, albeit still possible up to a maximum tolerable threshold usually indicated to be in a range between 50ms to 80ms [9], over which performance is impractical. The ultimate applicable threshold depends on factors such as skill, tempo, rhythm, and genre [4] or the application of special latency-coping performance strategies. Once the latency of a connection is assessed and minimized [10], the remaining available "latency budget" can be used by the network engineers to improve audio quality through longer buffers, apply jitter management, or employ higher streaming bitrates. Other effects such as audio auralization and spatialization are also occasionally applied to improve immersiveness and realism [11], [12], provided the added latency falls within an established tolerance level.

Literature on multimedia displays has consistently established that visual contact is a crucial element for fostering a sense of *presence* for musicians, and is at times more impactful than acoustic cues for the mutual understanding of expressive intentions [13]. Most commercially available tools for distributed performance support video streaming capabilities to enhance the connection experience [5]. However, video streams typically require higher bandwidths and compression latency than audio, leading to a higher overall transmission latency, sometimes in the order of hundreds of milliseconds, resulting in an out-of-sync rendering. In some cases, distributed musicians disregard the video feed when looking for tempo synchrony. Still, some scenarios allow for the insertion of additional latency in the audio buffer to re-synchronize with the late video stream [14], whenever that option is artistically viable [15].

When it comes to visual media, an alternative to video that has been explored in NMPs is motion-capture (mo-cap) data. Mo-cap is a format friendly to interactive XR environments, which has been indicated as a better visual tool in NMPs than standard video [16]. Mo-cap systems can transmit small data loads consisting of three-dimensional point coordinates representing human movement data. The capture stage necessitates the use of fixed tracking camera systems that follow the movement of trackers placed on special suits worn by participants [17]. The acquired data can then be live-streamed and interpreted at the receiver node to recreate a digital avatar through a graphics engine. This approach minimizes video delay, allowing for smooth visual interactions between participants thus being capable of improving the interactive visual experience and, by proxy, the musical collaboration. Several studies have started to explore the use of avatars in NMPs [18], [19] in combined music environments using custom combinations of multiple software that separately handle and process each data stream. To achieve virtual *copresence*, or the feeling of "being together" in a telematic space, it is desirable to place virtual avatars within a shared "realistic" auditory space [6], potentially paired with a cohesive acoustic environment coherent with the visual display, like a digital twin of a performance space [20].

III. Framework overview

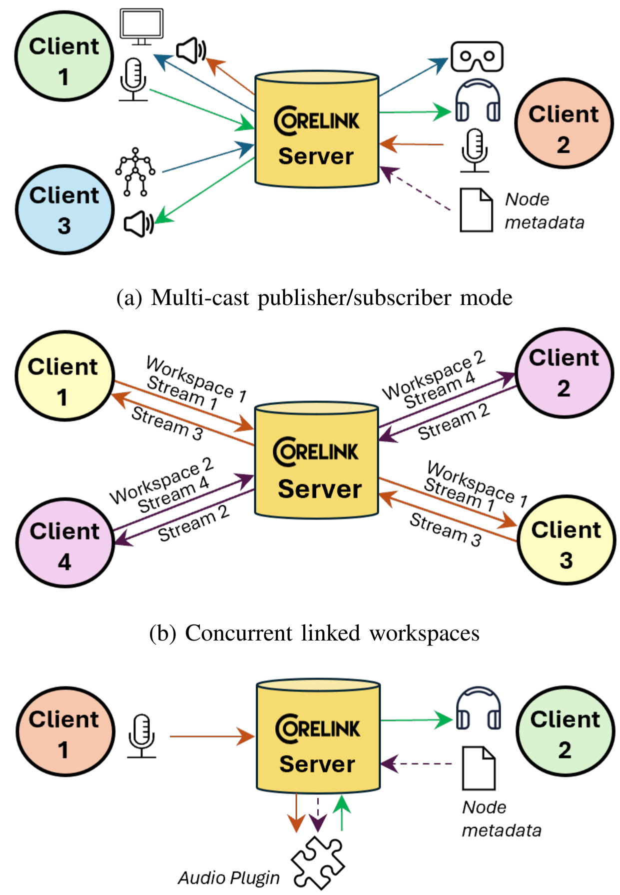
This section describes the framework of the *Holodeck* project [21], an inter-lab collaborative effort at *New York University* (NYU), supported by a distributed-stream management platform, *Corelink* [22]. This research and development framework serves as an important context for the considerations that went into the development of immersive NMP experiences. The combination of Corelink and Holodeck is used to empower multi-node adaptive immersive experiences, including network performances on arbitrary node architectures.

A. The Holodeck framework

The NYU Holodeck [21] is a collaborative project funded by the National Science Foundation aimed at creating an immersive virtual and physical research environment. This project focuses on creating an immersive and collaborative infrastructure equipped with advanced tools for national and international research, academic exploration, and creative output. The initiative's goal is to prototype the transition from conventional projection systems to functional augmented reality (AR), mixed reality (MR), and virtual reality (VR) simulations. This includes the integration of visual, audio, and physical components, along with technologies that enhance interactions among humans, agents, and robots.

In its NYU implementation, the Holodeck consists of a multi-room and multi-user augmented and interactive display built on top of a local ultra-low latency network that allows for an arbitrary number of general data streams to be distributed across the network. Each room in the network is capable of capturing data, transmitting through a dedicated server (Corelink), and rendering requested data that is adapted according to the local configurations. The consortium of participating laboratories involves facilities and experts from various disciplines, including computer science, simulation environments, urban planning, health informatics, digital media, and educational technology.

Several types of sensors are deployed in each connected laboratory room that belongs to the network. Cameras and motion tracking rigs are used to record video and body motion data, studio-quality microphones of different kinds capture high-quality low-noise audio for music and speech, and spatial audio microphones provide a soundfield capture (i.e. Ambisonics) of the room ambience. For reproduction, several options



(a) Multi-cast publisher/subscriber mode

(b) Concurrent linked workspaces

(c) Example of a Corelink architecture for an Audio Plugin

Fig. 1: Overview of Corelink functionality

are available. The incoming audio streams may be encoded in several formats and decoded to multichannel speaker arrays, or open-back headphones. This type of headphones allows for semi-transparent hearing which facilitates the interaction between local musicians during the performance. In regards to video and graphics, the displays can take both a two-dimensional form like a TV screen or a projection, or an immersive visual environment rendered on VR, AR, or MR headsets. For reaching higher levels of immersive cohesion (especially in mixed- or augmented-reality modes), most laboratories possess digital twin visual assets of their spaces and local acoustic measurement data. These can be used to adapt incoming streams to a locally characterized virtual audiovisual environment.

By developing a sophisticated supercomputing infrastructure, the project aspires to provide researchers and students with powerful tools for new knowledge creation and the application of integrated, multimodal techniques. Within this structure, the Holodeck is designed to be a flexible and reconfigurable system, capable of capturing comprehensive behavioral, physiological, and cognitive data, ultimately enabling detailed exploration of questions related to virtual environments, telepresence, and collaborative interactions.

B. Corelink data exchange system

Corelink [22] was initially created to materialize Holodeck's vision of a fully immersive and collaborative virtual research environment. The Corelink distributed stream management platform enables users to define, transmit, process, and receive real-time data in almost any format, including audio, motion capture, video, floorplans, maps, facial capture, virtual reality tracking, and more. Corelink handles all the routing required to establish a pipeline, broadcast, multicast, or any other network of connection possibilities (see Fig. 1). As a system, it can be classified as a data-agnostic and platform-neutral data exchange real-time protocol, built to facilitate low-latency data transport between a large number of senders and receivers. This is possible due to Corelink's dynamic streaming via a high-level routing abstraction and a high-performance publisher/subscriber data relaying architecture.

In essence, a Corelink connection consists of two main components: a server and a client. The Corelink server (manageable via API) acts as a centralized broker that manages authentication and access control information via a control channel and relays the user's data via a data channel. The control channel is established on the Secured WebSocket protocol [23] which runs reliably on top of the Transmission Control Protocol (TCP) and securely on top of Transport Layer Security (TLS). The data channel currently supports the usage of several protocols like the Transmission Control Protocol (TCP), User Datagram Protocol (UDP), and WebSocketSecure (WSS). No encryption scheme is used on the user data for TCP and UDP to minimize data transfer latencies. WSS uses encryption for use cases where encryption is desired.

Via Corelink's stream type specification, any number of data streams of various data types can be multiplexed and demultiplexed within a single internet application. This is particularly useful when virtual performances have multi-modality requirements. For example, audio capture and playback devices can be synchronized using a sample rate header, audio and video streams can be synchronized using a timestamping header, and motion-capture data can be coordinated with a cartesian values header. The Corelink routing protocol supports motion capture (mo-cap) data, which is crucial for immersive NMPs. The mo-cap "skeleton" data stream is captured by local tracking cameras and packetized into a Corelink transmission, then retrieved by the client node where is routed to a 3D graphics game for visual interpreting.

Corelink is ideal for the development of Networked Music Performance technology because there is no constraint on the data format. Up to 32768 bytes can be used for metadata or header and up to 65539 bytes can be used for the data, although this can vary depending on the chosen lower layer protocol. Custom plugins assist users in processing streams in real time. Examples include video transcoders, audio spatialization functions, and data encoding, but users can also create their own plugins and processing pipelines and manage them via workspace grouping. The Corelink project comes with a set of examples built on different clients such as C++,

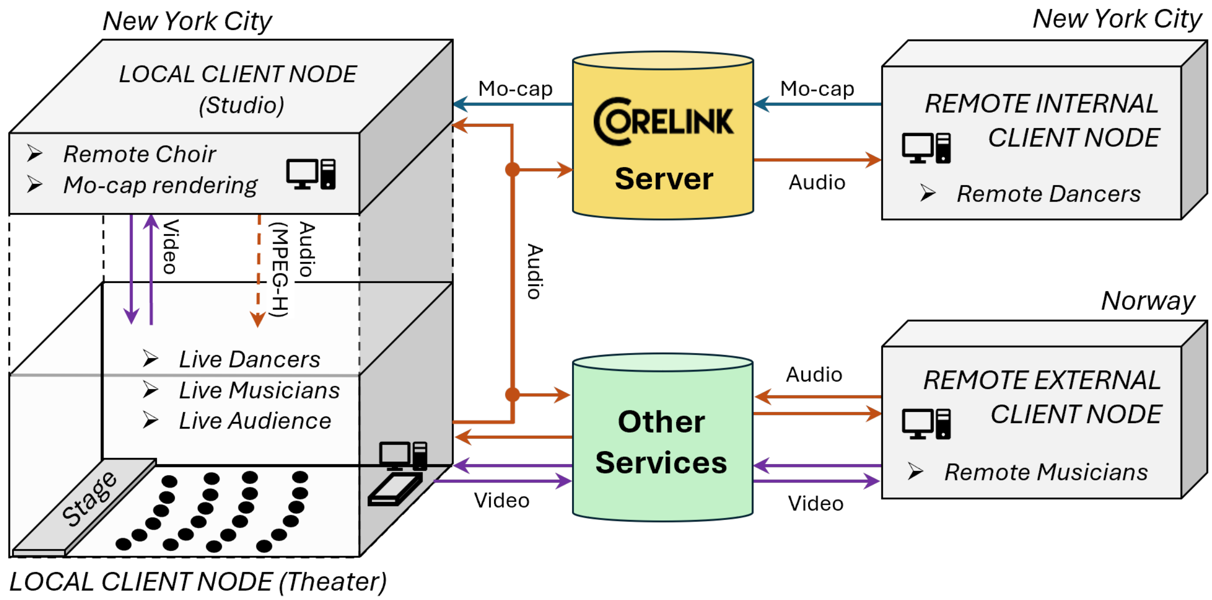


Fig. 2: Sketch of the Holodeck Concert: *Maps to the Stars*. The arrangement involved three nodes on an internal university network and an external node exchanging multimedia data.

JavaScript, and Python. The examples include group chat, iCreate robot remote control, video streaming on Web APIs, and audio streaming through JUCE processing. These example applications work out of the box and can be used to build multi-modal virtual performances.

C. *Previous work on the platform*

In [11] the authors previously piloted the application of a small-scale "hybrid" mixed-reality NMP supported by Holodeck and Corelink, between an exhibition node with live performers and audience, and a remote node with a number of performers (potentially pre-recorded) captured with motion-tracking and spot microphones. The experience is defined as "hybrid" due to the mix of a VR visual display (on a headset worn by the audience) showing the performer's avatars, and a MR auditory display, where the real and virtual sound of the collaborative distributed ensemble mixes together[2]. In this setup, the live performer acts as a "follower" in a "leader-follower" interaction paradigm [15] allowing the Corelink server to synchronize streams and render the signals without concern for latency issues.

To enable an immersive experience, the exhibition node's characterization data (i.e. metadata containing acoustic measurements and digital twin assets, see Fig. 1a and 1c) is shared with either a local machine or the Corelink server to process the incoming signals and adapt them to the local environment. The remote audio streams are thus auralized to include synthetic reconstructions of the early reflections and late reverberation derived from locally measured *Spatial Room Impulse Responses* [24]. Via game engine, the auditory scene is then arranged and spatialized according to the audience's perspective (orientation and positioning) in the exhibition space. The visual rendering is processed to render the live avatars of the locally present performers, and the streamed avatars of the remote nodes, within the digital twin environment of the local room. This pairing results in a cohesive audiovisual interactive display leading to a high degree of *immersion* and *realism* for the audience.

IV. Large-scale Distributed Multimedia Performances

The authors present a series of large-scale multimedia distributed performance concerts, that were designed to explore the challenges regarding the implementation of Corelink within the Holodeck paradigm. In the long term, these case studies are part of the research effort to open new possibilities for artistic collaboration and innovation whenever multiple music and dance artists are connected across complex spatial arrangements. These events piloted and established the groundwork for the creation of a novel research scenario, which enables studies otherwise impossible in traditional settings. In this scenario, contributions can be made towards the development of new musical interaction strategies and production methodologies that improve the quality and accessibility of remote collaborations, thereby expanding the reach and impact of musical and artistic expression. The ultimate goal of this research stream is to develop viable compelling improvements in immersive NMPs for both audiences and musicians and promote the use-case of remote collaborations.

This section describes the design and implementation of the *Holodeck Concerts*, two special multimedia distributed concert

[2]This setup is further discussed in a companion publication submitted by the authors at the same conference: *Locally Adapted Immersive Environments for Distributed Music Performances in Mixed Reality*

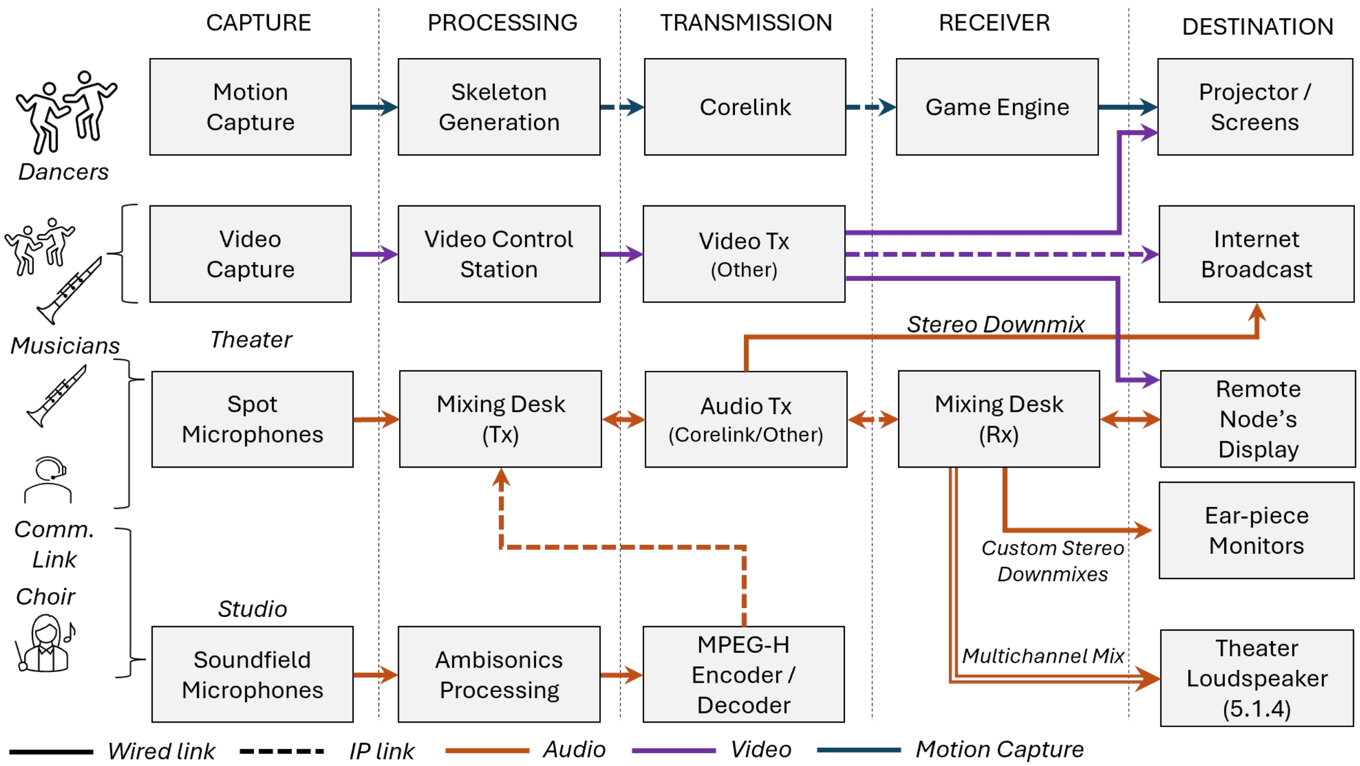


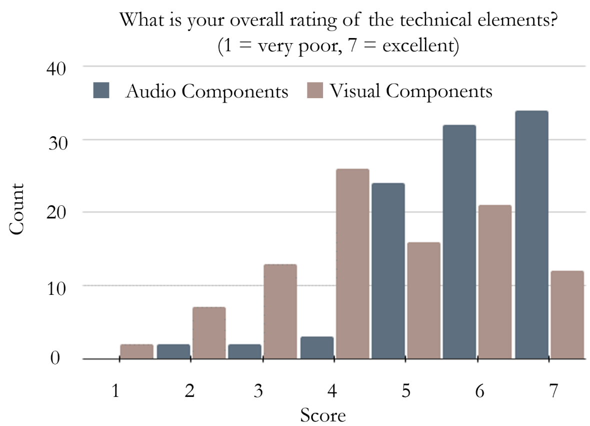
Fig. 3: Flow map of all media elements at the Holodeck concert *Ozark Henry, Maps to the Stars*.

performance events that occurred at *New York University*: *Ozark Henry on the Holodeck: Maps to the Stars* (October 2018), in collaboration with the Norwegian University of Science and Technology and other industry partners, and *Concert V* at the *Turn UP* music festival (March 2023) in collaboration with the University of Arizona [25]. These ideas were facilitated by the availability of an internal low-latency network infrastructure, capable of transmitting below 5ms. The integration of multimedia data exchange enhances the richness of the performance, allowing for a seamless blend of audio, visual, and motion data that can create more immersive and engaging experiences for audiences.

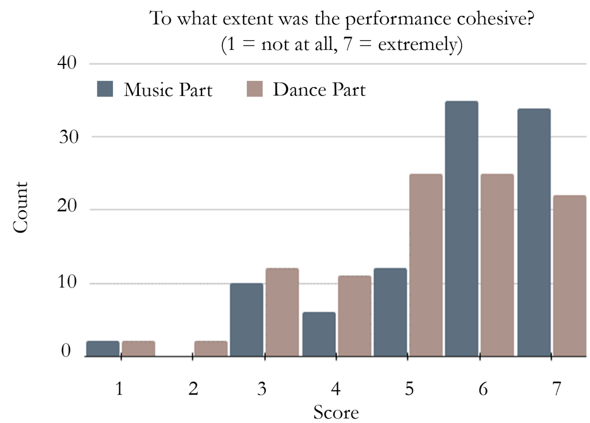
A. Concert: Ozark Henry on the Holodeck, Maps to the Stars

This distributed concert event ran in New York City in 2018 and featured live performances by international gold and platinum *Sony Music* recording artist Ozark Henry. The event served as an exploration into the creative application of the Holodeck, including Immersive sound technologies such as Ambisonics, live MPEG-H broadcast, multi-channel immersive sound, motion-captured dancers, and international collaborating artists. The spatial topology was designed as a four-node network created as a mix of intra-university nodes at NYU, connected to the local low-latency network infrastructure, and external nodes (specifically, located in Trondheim, Norway) connected over international internet links. Fig. 2 shows a sketch of the concert interactions map. The resulting setup is classifiable as a combination of a mesh- and a star-topology networks. While the internal nodes exchanged some data with each other, the theater node (co-located with the audience) acted as a reference "gravity" center, in which a link was established with all other local or remote nodes. Video and audio streams were also sent to a public internet live broadcast with an audio downmix created from the theater mixing desk. Fig. 3 shows a detailed overview of all interconnected elements.

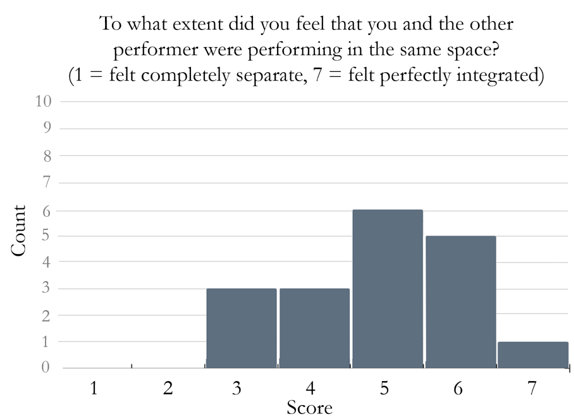
A theater location served as the primary node where the audience observed the live performers and where audio and video data were streamed for the stage display consisting of projectors and loudspeakers. Audio streams incoming from all of the nodes were mixed in the theater on the local console. There were several mixes created for different purposes. The main mix was created for the audience in the room using a 5.1.4 loudspeaker setup. Separate custom mixes were created for each of the musicians and dancers on the stage for in-ear monitoring. The set of separate stereo streams was mixed and sent to each of the remote nodes, to allow remote artists to sync to the lead musician located on the theater stage. The remote international node (Norway) was reached through standard NMP software tools to exchange audio channels and live video streams with the theater (latency $> 30ms$), while Corelink was used with the nodes connected to the internal IP infrastructure (latency $< 30ms$). The internal nodes are here classified as a "remote" mo-cap studio node enabling a dance performance, and a "local" music recording studio node, where a choir was present. The mocap studio was connected to the theater



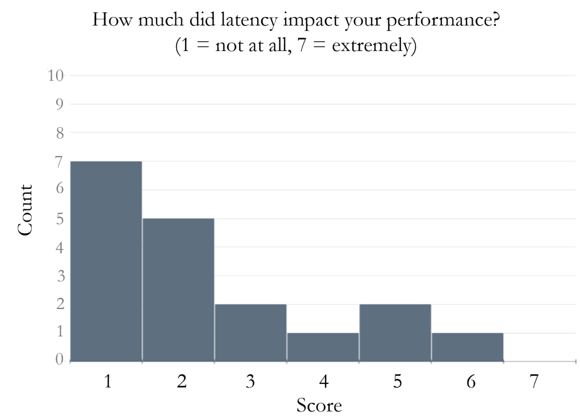
(a) Distributions of overall ratings on 7-point scale



(b) Distributions of artistic cohesion ratings on 7-point scale



(c) Distributions of copresence ratings on 7-point scale



(d) Latency impact on performance assessment on 7-point scale

Fig. 4: Post-event questionnaire distributions for audience $(N = 100)$ (top row), and performers $(N = 18)$ (bottom row)

Question	Scale	Mean	SD
Audience Questions $(N = 100)$			
How was your overall experience?	(1) Very Poor - (7) Excellent	5.60	1.15
What is your overall rating of all Audio components?	(1) Very Poor - (7) Excellent	5.90	1.10
What is your overall rating of all Visual components?	(1) Very Poor - (7) Excellent	4.63	1.53
Did you feel as though the choir was physically present in the theater?	(1) Not Present at all - (7) Extremely Present	5.45	1.48
Was the musical performance between stage and remote nodes cohesive?	(1) Not cohesive at all - (7) Extremely Cohesive	5.70	1.42
Was the dancing performance between stage and remote nodes cohesive?	(1) Not cohesive at all - (7) Extremely Cohesive	5.20	1.49
How did the technology impact the artistic material?	(1) Made it much worse - (7) Made it much better	5.43	1.21
Performer Questions $(N = 18)$			
For how many hours did you rehearse?	Continuous Numerical	6.08	5.35
For how many hours did you rehearse with a remote connection?	Continuous Numerical	1.12	2.02
How much did latency impact your performance?	(1) No impact at all - (7) Major impact	2.38	1.61
Did you feel as though performing in the same space?	(1) Completely separate - (7) Perfectly Integrated	4.88	1.18
Did the monitoring system impact your ability to feel immersed?	(1) Negative impact - (7) Positive Impact	4.22	1.93
As an artist, how enjoyable was the experience?	(1) Not enjoyable at all - (7) Excellently Enjoyable	5.38	0.97

TABLE I: Audience and Performers Questionnaire for the Holodeck Concert: *Ozark Henry, Maps to the stars* through Corelink to receive audio buffers of the performance mix. The received sound was played back locally to a motion-tracked dancer, the response movement data was transmitted back through Corelink towards the recording studio. The recording studio served as a pass-through server in which mo-cap data was rendered into humanoid avatars placed in a game engine environment. Being in the same building of the theater, where analog data channels were present, the recording studio was able to send a video capture of the rendered game engine scene onto the theater with virtually no delay, where it was finally projected to the audience.

The recording studio also hosted a live remote choir that received the theater audio through analog lines. The choir audio was captured using two multichannel spherical microphones placed in front of the ensemble to capture the sound in Ambisonics spatial format. Two industry partners, *THX* and *Qualcomm Inc.*, collaborated on the event by implementing an independent setup of a real-time MPEG-H encoder (Ateme live encoder), decoder, and renderer for spatial audio over loudspeakers. MPEG-H is an advanced audio coding standard designed to deliver high-quality immersive sound experiences, including 3D audio and interactive audio features, for a wide range of devices and applications [26]. The system was set up to encode and transmit Ambisonics soundfield audio recorded from the recording studio node and send it towards the mixing desk at the theater node, where the decoded audio was rendered for a 5.1.4 loudspeaker configuration using a spatial panning plugin.

Post-event survey: At the end of the event performance, the audience present in the theater ($N = 100$) was given a customized survey with Likert-type questions to fill out with instructions to provide various types of ratings and comments on the event as a whole. The audience sample was composed of 60% audio professionals, of which they had an average of 12.83 years of experience in their field (SD = 9.74). Fig. 4 shows the distribution of overall ratings of audio vs. visual components, and cohesion ratings of music vs. dance performance. For a full list of the questions asked please refer to Table. I. The captured results show how the visual components of the performance scored lower than the audio components in both overall rating (Fig. 4a) and performance *cohesiveness* (Fig. 4b). Free form feedback highlighted general satisfaction with the concert and its collaborative technology showcase, with some opinions for improvements. Several comments on the visual display being *"a bit dark and uncolored"*, *"avatar figures not coherent with the musical theme"* and similar remarks not related to the signal stability or latency but to the visual scene artistic quality and other brightness-related properties ($N = 13/100$). In regards to audio, a number of audience members ($N = 5/100$) described the experience of the spatially rendered choir to be slightly uneven based on the seating, suggesting that *"more mix should be sent to the high channels"* and that *"more speakers would have enhanced the choir and improved diffusion."* Sparse comments suggest that people in center seats had higher ratings of the choir, compared to side seating, where mentions of *"uneven loudness between choir and music"* were also reported ($N = 3/100$). A few audience members ($N = 5/100$) had comments on the excessive media latency of the visual components.

Some of the musicians present at the local NYC nodes (thus excluding the remote node in Norway, for which the scores are expected to be lower) were also surveyed and asked to evaluate their experience ($N = 18$) of which 14 performed on stage, and 4 reported previous experience with NMPs. By cross-pairing the survey results shown in Fig. 4, showing the copresence rating (i.e. *"feeling of being together in the same space"*, Fig. 4c [6]), with free-form feedback, there are indications that performers generally experienced different degrees of *copresence* according to their role in the concert and the amount of rehearsal in remote environments. More specifically, informal free-form feedback taken from the performers revealed that those in the studio felt *"more disjointed"* than performers on stage and that the opportunity to rehearse was crucial for the success and comfort of the distributed connection and musical approach. Similarly, the factor of latency impacted the self-evaluation of performance differently based on node location, affecting the studio musicians more than those at the stage, although generally a low impact was reported (Fig. 4d).

B. Concert: TURN UP Multimedia Festival

The TURN UP Multimedia Festival was a 3-day festival that featured 5 different concerts and took place in March 2023 [25]. This case study focuses on a collaboration between the *University of Arizona* (UA) and *New York University* (NYU) to do Networked Music Performances. Specifically, four total performances incorporated Corelink applications for real-time bidirectional audio streaming and unidirectional mo-cap streaming. This event served as an opportunity to demonstrate the feasibility of Corelink on long-distance networks, test the technical limits, and deploy user-friendly applications built on top of the Corelink library. No user experience audience/performer survey was collected for this event.

The two Corelink applications used for this concert were the *Corelink Audio plugin* and the *Corelink AR/VR system* for mo-cap streaming. Differently from the earlier events, Corelink was here tested and evaluated "in the field", outside of the controlled internal low-latency infrastructure present at NYU, but on a long-distance network. This led to higher network latencies and jitter that were accounted for by performers in the choice of the artistic style employed, leading to a less tempo-critical, "ambient", style of performance, or unidirectional "leader-follower" pieces [15] that would adopt a less interactive style of playing in favor of time synchronization at the audience (follower) node.

In this setup (Fig. 5), the audience was based in a theater at the UA node, where musicians and other artists performed on a stage. At NYU, two nodes on the NYU enterprise network (NYU-NET) connected to the UA theater and audience through Corelink. The NYU nodes consisted of a recording studio for musicians and a dance studio for motion-tracked dancers. The recording studio hosted a mix of singers

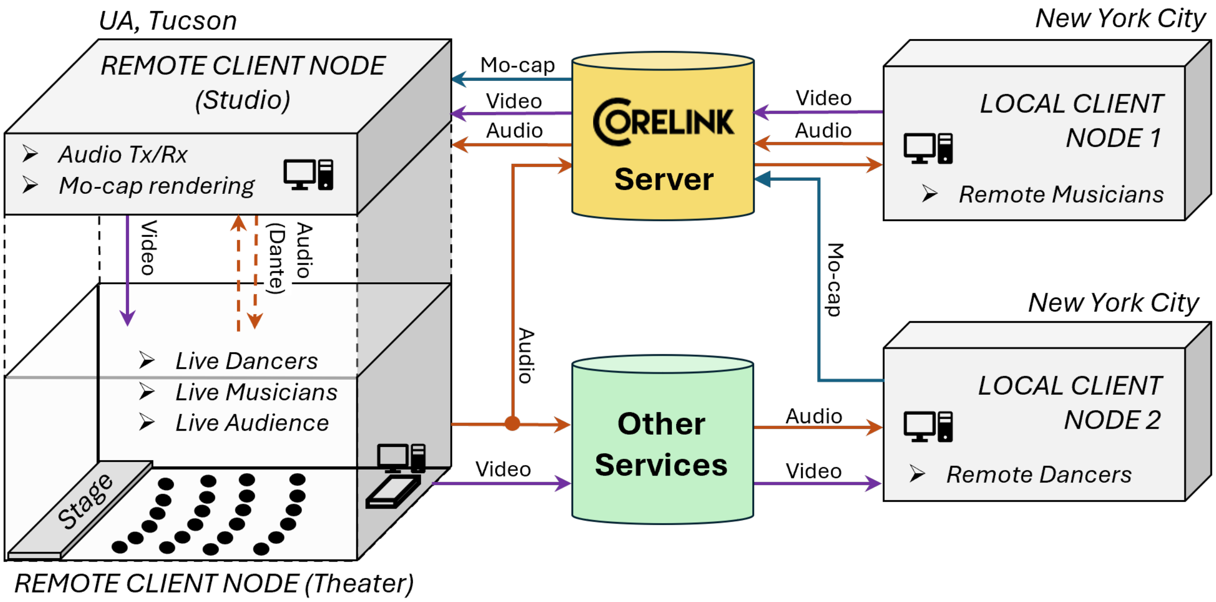


Fig. 5: Sketch of the Holodeck Concert at the *TURN UP Multimedia Festival*. The arrangement involved three nodes across two universities (NYU / UA)

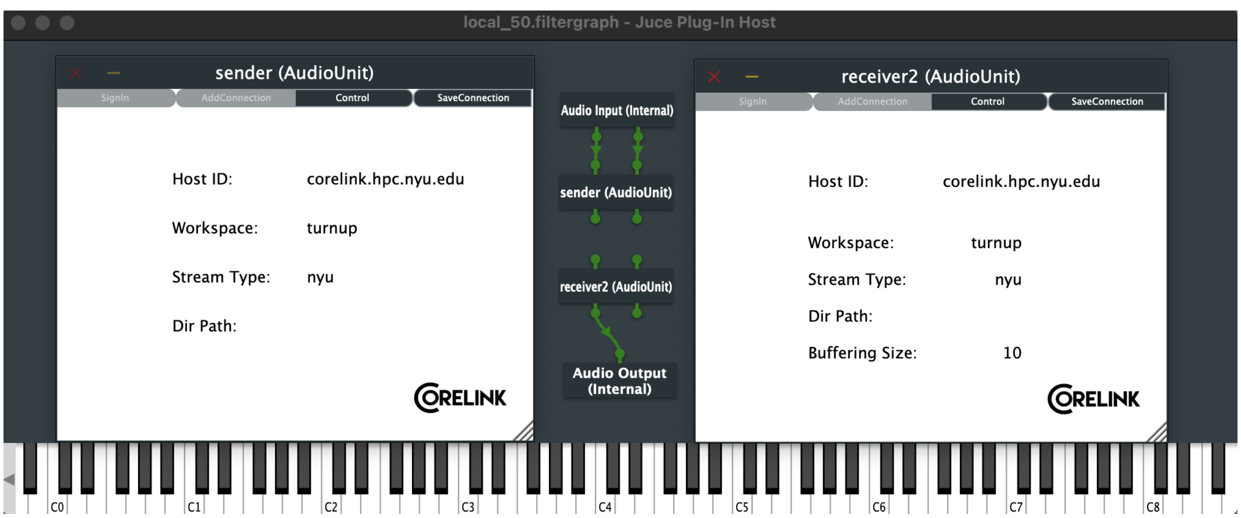


Fig. 6: Corelink Audio JUCE plugins for a sender and a receiver instance

and musicians bidirectionally receiving and transmitting audio channels towards the UA theater. On receiving captured audio data of musicians performing on the NYU site, the UA node routed the received audio from the studio booth to the stage's playback speakers via the Dante protocol. The dance studio unilaterally transmitted skeleton data, which was rendered on-location via a 3D graphics engine, and received audio from the theater. The Corelink server was hosted on the NYU High Speed Research Network cluster, accessible to all Corelink plugin instances. Because every Corelink data stream is relayed by the Corelink server, every node in the Corelink network is interconnected in a star topology.

In this implementation, the Corelink network framework library was paired with the JUCE framework to enable easy-to-use plugins, denominated as *Corelink Audio*. Through this integration, Corelink can run in different host environments as AU or VST plugin and has the ability to easily integrate with other audio or non-audio data streams with minimal networking setup. JUCE *AudioPluginHost* is a minimal plugin host application that can manage audio interface configurations, spawn AU/VST plugin instances, and do real-time audio stream porting. Using this application, users can directly adjust the sample rates and audio buffer sizes supported by their connected audio interfaces. Furthermore, the plugin editor also provides the users with input fields to specify the Host ID (the internet address for the Corelink server), the requested workspace to subscribe to, the stream types, and the playout buffer size for managing network jitter. To achieve a bidirectional audio connection, a user-friendly sender plugin as well as a separate receiver plugin need to run on one or more machines at each node (Fig. 6), along with the JUCE AudioPluginHost utility.

V. Discussion

Corelink future improvements: In regards to Corelink, the roadmap, maintained by the NYU High-Speed Research Network (HSRN), involves the stabilization of the code-base and the simplification of the development tools. In the long term, Corelink may provide data recording and customizable data analysis and processing functionalities for uses beyond NMPs. During the TurnUP festival, a number of periodic clicking artifacts were observed, suggesting that there might have been clock synchronization issues between the audio interfaces at each node. To tackle this issue, audio samples can be stretched and compressed in response to buffer underruns or overruns introduced by resynchronization. Furthermore, Corelink exhibited larger latencies compared to previous events ($\approx 30ms$). This finding was mostly attributed to the use of large playout buffers to compensate for the jitter typical of long-distance networks. Even with low jitter amounts, variations in latency still require large buffers to avoid clicking artifacts and dropped packets. Other potential reasons that may cause higher latency could lie in the JUCE overhead behavior and the reliability of the Corelink Server performance.

To further reduce the transport latency, there are plans for refactoring the server code in *Rust* as opposed to the current Node JavaScript implementation due to its capabilities of scheduling high-performance concurrency while enforcing strict memory safety rules (rather than Javascript event-driven system). Measurements show that Rust can outperform JavaScript by 115 times and overpower Node.js's concurrency model by 14.5 times [27]. For the physical network layer, new protocols may add additional latency improvements, such as the Remote Direct Memory Access over Converged Ethernet (RoCE) [28]. For sending a standard 1024-byte message, RoCE can be as fast as 1.814 μs compared to 13.55 μs in native Ethernet [29]. Falcon, a reliable low-latency hardware transport [30], also has the potential to help HSRN optimize its high bandwidth and low latency requirements as it scales up to assist researchers in facilitating more applications such as the one used by the Turn Up Multimedia Festival concert for real-time audio and mocap streaming.

Survey: Some of the most evident outcomes of this preliminary pilot research are that audiences were more affected by visual factors rather than auditory and that the spatial immersive display was well received although improvements in the mix distribution have been suggested. It should be noted that the audience for this event was mostly audio critics and experts, opening the possibility that the nature of the audience may have affected evaluation of the performance. Regarding the performer questionnaires, there were informally communicated differences in the *presence* experience between studio and stage, suggesting that the studio experience could be improved by adding auralization to help this imbalance. It was also noted that rehearsal was beneficial for the performers' comfort. Crucial data is missing due to the inability of the remote node musicians to respond to the poll (for which scores are expected to be lower). More case studies need to be explored before developing research hypotheses and establishing baseline priors for *Bayesian-type* analysis.

The choice of questions presented to audience and performers was not based on a predefined questionnaire template, but was drafted as a "large-scale pilot" customized by the authors for capturing specific insights on the particular event. The authors intend to improve and generalize the multimedia audience-performer questionnaire in future iterations by drawing and adapting from validated sources that have been proposed in the years since the concert event. Some potential sources are the System Usability Scale [31], the User Experience Questionnaire [32], and other recent NMP-specific formats [33], [34]. Overall, the discussion around subjective responses from audiences can help to establish an evaluation benchmark over which improvements can be tracked over time. Our presented scores are thus useful considerations for investigating major trends and distributions of mean opinion scores on large-scale NMPs.

VI. Conclusions

This paper describes the *Holodeck* infrastructure and the *Corelink* open-source platform, implemented at New York University, and their application in network-based distributed performances (NMPs). These systems facilitate a multi-room network capable of flexibly exchanging multimedia audio-visual data, which can be rendered at each receiving node according to its specific capabilities and requirements. This unitary and user-friendly platform facilitates research in both NMPs and the Musical Metaverse [7] by supporting new forms of "Musical XR" thanks to interactive displays and multimedia data streams while maintaining core data transmission and broadcasting functionality adaptable to arbitrary network topologies and data stream types.

To demonstrate the goals and current capabilities of the system, several case studies of large-scale network music performances were presented. These performances tested the transmission of multichannel audio and motion-capture data over different node configurations, contributing to the development of innovative methodologies and artistic practices in multimedia performances. Additionally, the system was utilized for piloting research on subjective impressions of NMPs, gathering preliminary data on the quality of experience from both audiences and musicians. The findings from these pilot events underline the potential of the Holodeck and Corelink platforms to enhance telepresence and realism in remote collaborations, promoting further exploration and refinement of distributed performance technologies.

Future Work: Having set a model infrastructure and experience-design process, the future work around Holodeck and Corelink will focus on the exploration of its boundaries and how they relate to artistic musical practices. There are several open research questions on how interactive multimedia displays in NMPs are affected by elements such as different media combinations, number of nodes, hierarchical musical organizations, multimodal rendering asymmetries, artistic goals and techniques, immersive displays, and more. The Corelink and Holodeck platforms aim to provide a canvas over which this research is made possible.

In the long term, the holy grail of NMP and immersive media engineering is to democratize existing tools and move towards the possibility of transferring the experience to a mobile ecosystem [35], available at a low cost to the public. Having flexible professional tools for deploying innovative environments for remote performance is paramount for the development of a purposeful artistic practice, that sees NMPs not just as a contingency plan for forced remote interactions or as an experimental setting, but also as a viable collaborative platform for artistic education, rehearsals and concerts.

ACKNOWLEDGMENT

The authors would like to acknowledge Prof. Thomas Beyer who led the creative component of the Holodeck concerts, Prof. Paul Geluso who directed the stage engineering, and Michael Oikonomidis who ran the studio facilities. The authors also thank all the students and research assistants who participated in these events across all the universities involved, the Corelink development team, the performers, the sound engineers, the NYU Holodeck Consortium, and the audiences.

REFERENCES

[1] K. E. Onderdijk, F. Acar, and E. Van Dyck, "Impact of lockdown measures on joint music making: playing online and physically together," *Frontiers in Psychology*, vol. 12, no. 5, 2021.

[2] L. Turchet, R. Hamilton, and A. Çamci, "Music in extended realities," *IEEE Access*, vol. 9, pp. 15 810–15 832, 2021.

[3] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of musical things: Vision and challenges," *IEEE Access*, vol. 6, pp. 61 994–62 017, 2018.

[4] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016.

[5] M. Bosi, A. Servetti, C. Chafe, and C. Rottondi, "Experiencing remote classical music performance over long distance: A Jacktrip concert between two continents during the pandemic," *Journal of the Audio Engineering Society*, vol. 69, no. 12, pp. 934–945, 2021.

[6] M. F. Schober, "Virtual environments for creative work in collaborative music-making," *Virtual Reality*, vol. 10, no. 2, pp. 85–94, 2006.

[7] L. Turchet, "Musical metaverse: vision, opportunities, and challenges," *Personal and Ubiquitous Computing*, vol. 27, no. 5, pp. 1811–1827, 2023.

[8] A. Boem and L. Turchet, "Musical metaverse playgrounds: exploring the design of shared virtual sonic experiences on web browsers," in *2023 4th International Symposium on the Internet of Sounds*, 2023, pp. 1–9.

[9] C. Chafe, J.-P. Caceres, and M. Gurevich, "Effect of temporal separation on synchronization in rhythmic performance," *Perception*, vol. 39, no. 7, pp. 982–992, 2010.

[10] R. Hupke, S. Sridhar, A. Genovese, M. Nophut, S. Preihs, T. Beyer, A. Roginska, and J. Peissig, "A latency measurement method for networked music performances," in *Audio Engineering Society Convention 147*. Audio Engineering Society, 2019.

[11] A. Genovese, M. Gospodarek, and A. Roginska, "Mixed Realities: a live collaborative musical performance," in *5th International Conference on Spatial Audio ICSA*, Ilmenau, Germany, 2019.
audio system," *Journal of Audio Engineering Society*, vol. 69, no. 12, pp. 924–933, 2021.

[12] P. Cairns, H. Daffern, and G. Kearney, "Parametric evaluation of ensemble vocal performance using an immersive network music performance

[13] J. W. Davidson, "Visual perception of performance manner in the movements of solo musicians," *Psychology of Music*, vol. 21, no. 2, pp. 103–113, 1993.

[14] C. Drioli, C. Allocchio, and N. Buso, "Networked performances and natural interaction via LOLA: Low latency high quality A/V streaming system," in *Information Technologies for Performing Arts, Media Access, and Entertainment*, P. Nesi and R. Santucci, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 240–250.

[15] A. Carôt and C. Werner, "Fundamentals and principles of musical telepresence," *Journal of Science and Technology of the Arts*, vol. 1, no. 1, pp. 26–37, 2009.

[16] B. Loveridge, "Networked music performance in virtual reality: current perspectives," *Journal of Network Music and Arts*, vol. 2, no. 1, p. 2, 2020.

[17] C. Bui, A. Genovese, T. Bradley, and A. Roginska, "Multimodal immersive motion capture (MIMiC): A workflow for musical performance," in *Audio Engineering Society Convention 149*. Audio Engineering Society, 2020.

[18] A. Hunt, H. Daffern, and G. Kearney, "Avatar representation in extended reality for immersive networked music performance," in *AES International Conference on Spatial and Immersive Audio*, Huddersfield, UK, 2023. [Online]. Available: https://unity.com/

[19] R. Hupke, S. Preihs, and J. Peissig, "Immersive room extension environment for networked music performance," in *153rd Audio Engineering Society Convention*, New York, USA, 2022.

[20] A. F. Genovese, *Acoustics and Copresence: towards effective auditory virtual environments for distributed music performances.* New York University. Ph.D. Thesis., 2023.

[21] (2017) Holodeck - NYU. [Online]. Available: https://holodeck.nyu.edu//

[22] (2019) Corelink: Real time networking for research. [Online]. Available: https://corelink.hsrn.nyu.edu/

[23] A. Lombardi, *WebSocket: lightweight client-server communications*. O'Reilly Media, Inc., 2015.

[24] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *Journal of the Audio Engineering Society*, vol. 61, no. 1/2, pp. 17–28, 2013.

[25] (2024) Turn Up Multimedia festival 2023. [Online]. Available: https://turnupfestival.weebly.com/turn-up-2023.html

[26] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "Mpeg-h 3d audio—the new standard for coding of immersive spatial audio," *IEEE Journal of selected topics in signal processing*, vol. 9, no. 5, pp. 770–779, 2015.

[27] K.-I. D. Kyriakou and N. D. Tselikas, "Complementing javascript in high-performance node. js and web applications with rust and webassembly," *Electronics*, vol. 11, no. 19, p. 3217, 2022.

[28] G. Kaur and M. Bala, "Rdma over converged ethernet: A review," *International Journal of Advances in Engineering & Technology*, vol. 6, no. 4, p. 1890, 2013.

[29] L. A. Kachelmeier, F. V. Van Wig, and K. N. Erickson, "Comparison of high performance network options: Edr infiniband vs. 100gb rdma capable ethernet," Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), Tech. Rep., 2016.

[30] D. Lenoski and N. Dukkipati, Oct 2023. [Online]. Available: https://cloud.google.com/blog/topics/systems/introducing-falcon-a-reliable-low-latency-hardware-transport

[31] J. R. Lewis, "The system usability scale: past, present, and future," *International Journal of Human–Computer Interaction*, vol. 34, no. 7, pp. 577–590, 2018.

[32] M. Schrepp, A. Hinderks *et al.*, "Design and evaluation of a short version of the user experience questionnaire (ueq-s)," 2017.

[33] P. Cairns, A. Hunt, D. Johnston, J. Cooper, B. Lee, H. Daffern, and G. Kearney, "Evaluation of metaverse music performance with bbc maida vale recording studios," *Journal of the Audio Engineering Society*, pp. 313–325, 2023.

[34] cairns patrick, rudzki tomasz, cooper jacob, hunt anthony, steele kim, acosta martínez gerardo, chadwick andrew, daffern helena, and kearney gavin, "singer and audience evaluations of a networked immersive audio concert," *journal of the audio engineering society*, vol. 72, pp. 467–478, september 2024.

[35] L. Turchet, C. Rinaldi, C. Centofanti, L. Vignati, and C. Rottondi, "5g-enabled internet of musical things architectures for remote immersive musical practices," *IEEE Open Journal of the Communications Society*, 2024.