Toward an Inclusive Framework for Remote Musical Education and Practices

Original
Toward an Inclusive Framework for Remote Musical Education and Practices / Rottondi, Cristina; Sacchetto, Matteo; Severi, Leonardo; Bianco, Andrea. - In: IEEE ACCESS. - ISSN 2169-3536. - 12:(2024), pp. 173836-173849. [10.1109/access.2024.3501414]

Publisher:
Institute of Electrical and Electronics Engineers - IEEE

Published
DOI:10.1109/access.2024.3501414

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 19 October 2024, accepted 14 November 2024, date of publication 18 November 2024,
date of current version 29 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3501414

TOPICAL REVIEW

Toward an Inclusive Framework for Remote Musical Education and Practices

CRISTINA ROTTONDI, (Senior Member, IEEE),
MATTEO SACCHETTO, (Graduate Student Member, IEEE),
LEONARDO SEVERI, (Graduate Student Member, IEEE),
AND ANDREA BIANCO, (Senior Member, IEEE)
Department of Electronics and Telecommunications, Politecnico di Torino, 10129 Turin, Italy

Corresponding author: Cristina Rottondi (cristina.rottondi@polito.it)

This work was supported by the Musical Metaverse: An Inclusive Extended Reality Platform for Networked Musical Interactions Project funded by European Union (EU)-Next Generation EU within the Progetti di Rilevante Interesse Nazionale (PRIN) 2022 Program (D.D. 104-02/02/2022 Ministero dell' Università e della Ricerca) under Grant 2022CZWWKP. The work of Leonardo Severi was supported by the Ph.D. Program funded by the European Union in the framework of the Resiliency and Recovery Plan (RRP) within the NextGenerationEU Initiative.

ABSTRACT This paper provides an overview on inclusiveness in remote music education and networked music performances, highlighting issues and difficulties encountered by users with visual, auditory or motor disabilities when approaching such practices. Audio-visual tools, motion tracking systems, immersive audio and haptic feedback technologies that can be leveraged to enhance the accessibility of web-based platforms for online music teaching and remote performances are analyzed and classified. Moreover, the manuscript discusses the integration of such technological solutions in a networked music performance framework and their potential benefits for various categories of disabled users, describing some exemplary use-cases. Finally, open technical challenges for the practical realization and deployment of such a framework are identified, as well as future research directions towards its incorporation within the so-called Musical Metaverse.

INDEX TERMS Inclusiveness, networked music performances, remote education, online music teaching.

I. INTRODUCTION

Information and Communication Technologies (ICT) and Artificial Intelligence (AI) are introducing dramatic changes in the way pedagogical processes are conceived: distance education and computer-based learning have started proliferating in the last decade and are expected to further expand [1], fostered by novel, interactive and immersive approaches enabled by technologies such as Internet of Things (IoT), Virtual and Extended reality (VR/XR), and the Metaverse [2]. This implies a global rethinking of the fruition modalities for educational services.

There are several categories of subjects who would benefit from the long-term adoption of distance learning frameworks. Among those, students with visual/auditory/motor impairments or special needs, for whom traditional in-presence access to education is challenging or even precluded, represent a non-negligible share. According to [3], in 2020/21 15% of US public school students received special education services, for a total of 7.2 millions. Such amount comprises more than 1 million students diagnosed with health impairments resulting in limited strength, vitality, or alertness, around 700 thousands students with hearing impairments, and around 350 thousands students with orthopedic/visual impairments or deaf-blindness. The same source reports that 1% of students with special educational needs are homebound or confined in hospitals or separate residential facilities.

As of today, remote teaching heavily relies on videoconferencing technologies, which are mainly designed to maximize speech intelligibility while reducing the required transmission bit-rate, resulting in the adoption of low sampling frequencies and highly efficient compression codecs, at the price of causing voice signal distortions. On the other hand, the encoding/decoding processes add consistent

The associate editor coordinating the review of this manuscript and approving it for publication was Lei Shu.

latency to increase resiliency to packet losses [4]. Differently from from videoconferencing, in the application field of remote music teaching and Networked Music Performance (NMP) scenarios, where musicians play together in real-time from distance by exploiting audio streaming over wide area networks, specific requirements emerge, such as:

- strict latency guarantees, in the order of at most 30 ms [5], to ensure realistic performative conditions;
- high reliability, to ensure high-fidelity quality of the audio playout;
- commensurate computational capabilities for audio processing, mixing and rendering, to ensure scalability to large deployments (e.g., for music classes, ensembles, choirs or orchestras).

As a consequence, the applicability of off-the-shelf videoconferencing technologies is precluded in remote musical education and NMP scenarios. In such applications, audio codecs would worsen the perceived end-to-end delay and reduce the quality of the streamed audio signal due to their compressive effects. Moreover, an adequate backend infrastructure to support low-latency and high-fidelity audio mixing would be required. The aforementioned limitations clearly emerge in [6], which highlights the difficulties encountered by music schools in delivering group lessons and supporting rehearsal sessions using commercial e-learning technologies during the first lockdown enforced to mitigate the spreading of the SARS-CoV-2 pandemic.

Motivated by the fact that the above-mentioned shortcomings heavily hinder the widespread adoption of distance education of music-related subjects, and consequently their adoption by subjects with disabilities, the aim of this paper is threefold:

- offering a comprehensive overview of the educational and performative needs experienced by disabled users in the field of remote musical teaching and practices;
- identifying the technological solutions that could address such needs;
- discussing their integration in existing NMP frameworks to enhance the inclusiveness of their adoption, highlighting open challenges and future research directions.

The rest of the manuscript is organized as follows: Section II reviews accessibility and interaction requirements to enable fruition of remote musical education and practices by several categories of disabled users. It also discusses the main technological solutions for the realization of accessible human-machine interfaces tailored to musical performances. Section III overviews existing NMP applications and compares their design choices in terms of implementation and infrastructure. Section IV envisions an inclusive NMP framework and details its main building blocks. Some illustrative use cases are presented in Section V. Open technical challenges and future research directions involved in its realization are discussed in Sections VI and VII, respectively. Finally, VIII concludes the paper.

II. ACCESSIBLE HUMAN-MACHINE INTERFACES FOR MUSICAL EDUCATION

A. EDUCATIONAL AND PERFORMATIVE ISSUES EXPERIENCED BY DISABLED USERS IN MUSIC-RELATED FIELDS

Research on how to accommodate the needs of students with disabilities in the field of music education started more than four decades ago, as testified by two surveys, respectively, appeared in 2007 [15] and 2015 [16]. Such studies triggered further investigations on how music teaching can embrace inclusiveness and diversity [17]. In Table 1, we report some of the most widely recognized barriers that preclude or strongly limit the inclusion in musical education of subjects with various types of disabilities, as well as a number of resources available to educators to make musical practices accessible to students with special needs.

In the following subsections, we review the main technological solutions already adopted for in-presence inclusive musical education. In particular, in the past two decades the research areas referred to as “Musical Extended Reality” (Musical XR) [18] and Internet of the Musical Things (IoMusT) [19] have started being investigated in the scientific community. The former introduced a paradigm shift by disrupting traditional notions on musical interaction and enabling performers and audiences to interact musically with virtual/augmented objects, agents, and environments [20], also in the context of music education [21]. Conversely, the latter identifies computational devices integrated into physical objects (e.g., smart musical instruments or wearables with music-related purposes) devoted to generating or receiving musical content, interconnected via a telecommunication infrastructure. Such smart musical objects can be leveraged also for educational purposes. Finally, we complete the section by reviewing some literature-reported use-cases that testify the adoption of the presented technologies in inclusive music education trials.

B. HAPTIC FEEDBACK TECHNIQUES

A number of studies have investigated the conveyance of musical information through the sense of touch (see [22], [23] for a thorough overview). Tactile rendering setups normally focus on capturing a single music feature, e.g., rhythm, pitch, melody, timbre, and loudness.

The usage of musical haptic wearables to improve communication among performers sharing the same physical environment has been envisioned in [24]. The adoption of haptic feedbacks to enrich the emotive experience while listening to music has also been experimented in [25] and [26]. Furthermore, systems that transpose gestural cues into haptic feedbacks have proven to be useful to enrich musical perception and interplay for Blind and Visually Impaired (BVI) and Deaf and Hard of Hearing (DHH) performers [27].

Notably, the sensitivity of the tactile system is quantified in [28] in terms of frequency, intensity and temporal features. The study reports that, through haptic stimulation at the

TABLE 1. Overview of educational issues of disabled music students.

Type of Impairment	Issues	Countermeasures
Blind and Visually Impaired (BVI) students	Being music notation predominantly visual, it cannot be written, read or directly reproduced by the majority of BVI musicians, which heavily hinders their capability of conveying their creative intentions [7]–[9];	Availability of braille scores [10], modified musical instruments [11] and accessible music production interfaces [12].
Deaf and Hard of Hearing (DHH) students	DHH subjects experience difficulties in fine pitch discrimination, which often precludes practicing with instruments such as violin, cello or brass instruments, all of which require very precise intonation. Moreover, DHH musicians rely on visual cues to compensate for their reduced hearing capabilities, which may constrain their physical placement when playing in ensembles, bands, and orchestras [13].	Choice of instruments which easily convey vibrations to DHH players (e.g., guitar, electric bass). Strategic pairing of non-DHH to DHH performers to ensure dedicated guidance/cues [13]. Availability of modified musical instruments [11].
Mobility impaired students	Subjects with motor impairments often cannot play traditional musical instruments or use music production software due to lack of accessibility [14].	Wide offer of adapted musical repertoire (e.g., one-handed piano pieces) [10]. Availability of modified musical instruments [10].

arm, wrist, or hand, approximately 40 different frequency steps can be discriminated. Moreover, amplitude modulation sensitivity was found to be highest when using a carrier tone at 250 Hz. Finally, audio and haptic signals are perceived as simultaneous if the discrepancy between the haptic and audio signal onsets is within 25 ms.

In terms of bit-rate requirements, tactile feedback ranges from tens of to hundreds of kbps, depending on the type of stimuli being involved.

C. MOTION TRACKING TECHNIQUES

In the context of music education, several studies have already demonstrated the benefits of the usage of computer technologies for BVI people [29]. However, no scientific study has yet tackled the integration of motion tracking or haptic feedback functionalities in NMP systems to accommodate the special needs of disabled users. The exploitation of Kinect cameras for telerehabilitation has been investigated in [30] using a WebRTC-based communication protocol, though without integration of audio streaming functionalities. Kinect cameras for motion data acquisition have already been adopted to analyze gestures of music performers [31], [32].

Notably, from a networking standpoint, streaming pose data for avatar rendering rather than video frames significantly reduces transmission bit-rate requirements, as it only needs to convey the Cartesian position and rotation data of the represented joints. This results in savings of at least one order of magnitude compared to encoded video streaming, and potentially up to three orders of magnitude for unencoded video. Regarding motion-to-photon latency (i.e., the time it takes for a user's gesture to produce a visual change on a display), experimental values are below 40 ms [33], which is comparable to the delays introduced by a camera streaming compressed video frames. This would foster scalability in large deployments involving a high number of participants, (e.g., choirs or orchestras), bypassing the need to convey a dedicated video stream to each of them.

D. MUSIC VISUALIZATION TECHNIQUES

Visualization can also be exploited to present music content to subjects with auditory impairments: within the rich corpus of scientific literature dedicated to such a topic (see [34] for a comprehensive survey), a few studies have specifically targeted the DHH community. In paper [35], sounds generated by a violin are digitized and associated with movements of objects such as flowers and plants, with the goal of achieving interactive sound visualization to support music education for children with hearing impairments.

Paper [36] presents ViTune, a music visualization prototype that enhances the musical experiences of the DHH community thanks to the usage of an on-screen visualizer generating effects alongside music.

Paper [37] introduces BufferBeats, a toolkit for creating multimodal experiences for streaming services, aimed at improving accessibility to online music streaming by DHH subjects. The study is motivated by the presence of Digital Rights Management anti-piracy encryption, which often restricts the access to audio data which is required to create multimodal experiences for music streaming.

To the best of our knowledge, music visualization techniques have always been adopted in local settings or applied to pre-recorded streamed audio material only, whereas their integration in NMP frameworks for real-time musical practices has not yet been reported in the literature. However, since such techniques typically rely on local processing of audio content, from a networking perspective they do not require transmission of other data than audio, thus not incrementing the transmission burden of NMP systems.

E. IMMERSIVE AUDIO RENDERING

Immersive audio is rendered by means of sound diffusion systems aimed at offering to the listener a three-dimensional (3D) representation of acoustic sources, so that they can be perceived as coming from a specific spatial direction. For the specific use-case of NMP applications, the usage of headphones is predominant with respect to surround sound systems, thanks to their ease of portability and configuration, and lower cost. In such a scenario, binaural audio rendering is typically the adopted immersive audio solution. It relies on the rendering of acoustic cues such as interaural time/level differences and acoustic filtering (i.e., the spectral information that depends on the characteristics of the user's physical attributes such as the shape of ears, head, shoulders, torso). This rendering is achieved via

Head-Related Transfer Functions (HRTFs), which convey the directionality of a sound source to the listener's eardrum [38].

Spatial audio technology has already proved to enhance the sense of virtual imagery and spatialization for BVI users [39], [40], [41], [42]. In the context of musical practices, 3D audio rendering can be leveraged to recreate a shared acoustical scene at the premises of every remote participant. Therefore, it is expected to facilitate the participation to ensemble performances of BVI subjects, who could take advantage of the spatialized placement of sound sources to enhance their perception of the auditory cues necessary to synchronize their playing with that of the other performers. Similarly, DHH users could benefit of a personalized design of the acoustic scene rendered at their premises, tailored to their specific auditory impairment.

Though a few examples of immersive and multimodal environments specifically designed to permit fruition also by subjects with disabilities exist (see, e.g., [43]), no attempt to integrate immersive audio rendering systems in inclusive NMP frameworks has yet been reported in the literature. However, immersive audio settings for NMPs have started being investigated and assessed [44], [45], [46], [47].

Concerning bit-rate requirements, they depend on the desired sampling rate and amount of bits per sample, as well as on the number of audio channels to be supported. For example, an uncompressed mono audio stream at 48 kHz using 16 bits resolution accounts for 768 kbps. Such amount scales linearly with the number of audio channels. It follows that scenarios involving multiple participants and complex spatialized audio renderings may impose a significant burden in terms of data volumes to be transferred, potentially non compatible with up/downstream bit-rates supported by typical residential connectivity.

F. EXEMPLARY FIELD TRIALS

In in-presence music teaching, vibrotactile feedbacks have been already demonstrated to convey the concept of rhythm and of the spectrum of sound frequencies to BVI and DHH students [48], [49]. Moreover, motion tracking technologies in connection with sound and visual production have been experimented to ease musical expressions of subjects with mobility or cognitive impairments [50]. The potentialities connected to the adoption of virtual and augmented reality technologies for music teaching have been investigated in several studies, especially for primary education [51], focusing on mobile settings [52], and even including disabled students [21]. Notably, the combination of passive haptics with virtual reality has proved useful for music conduction education [53].

Focusing instead on on-line music education, several pilot studies conducted with traditional videoconferencing systems have been documented (see, e.g., [54]), especially during the Sars-CoV-2 pandemic period [55], [56]. However, as emerged in [57], sustainable technology integration in music classrooms is still far from being achieved at large scale, mainly due to barrier such as lack of digital literacy, non-affordability of equipment costs and scarce alignment with the instructors' pedagogical praxis. This outcome is in line with the general trend observed in e-learning-based education, as testified by numerous studies on its technological acceptance in various geographical areas (see, e.g., [58], [59], [60], [61]). Another issue is the economical sustainability of ICT-based educational solutions, for which dedicated pricing models and cost-benefits analyses must be devised, as recently investigated in [62], [63], and [64].

III. NMP APPLICATIONS AND BACKEND INFRASTRUCTURES

A. MAIN FUNCTIONALITIES OF NMP SYSTEMS

The core functionality of NMP systems is the support of professional-quality real-time audio streaming, i.e., with mouth-to-ear latency figures comparable to those achievable in in-presence settings (ideally 20-30 ms, which correspond to the time required by a sound wave to propagate in air to cover a distance of 8-10 m). Above such latency threshold, the musicians' delayed perception of the sounds produced by their remote counterpart(s) starts negatively impacting the musical interplay. Indeed, latency figures above 20 ms induce a tendency towards tempo deceleration, thus making the maintenance of a stable tempo more challenging [5].

Additional features typically integrated in NMP systems are:

- local audio manipulation capabilities (e.g., recording, reproduction of pre-recorded streams, adjustment of volumes, equalization levels and panning);
- local digital audio workstations;
- communication and interaction interfaces such as chats, user status tracking etc.;
- collaborative music composition/editing/production tools.

For what concerns video support, only a few NMP solutions integrate video streaming capabilities. Unfortunately, typical video acquisition and encoding/decoding delays significantly exceed the 30 ms latency threshold, which is incompatible with remote musical performances. Therefore, the offered video-stream is usually desynchronized w.r.t. the corresponding audio-stream, thus heavily hindering reliance on visual cues during real-time remote musical interactions [65]. Alternatively, video is streamed without being encoded (as in [66], [67]), which requires an adequate telecommunication infrastructure capable of supporting an uplink/downlink bit-rate of a few Gbps.

B. PEER-TO-PEER VS CLIENT-SERVER ARCHITECTURE

Depending on the implementation choices and on the scalability needs of the users, either client-server or peer-to-peer architectural solutions are adopted. Typically, peer-to-peer architectures offer lower mouth-to-ear latency than client-server ones, but can accommodate only a few musicians in the same session. Conversely, for larger deployments client-server approaches are preferred, at the price of added latency caused by the server operating as intermediate relay node. The scalability limitations introduced by peer-to-peer architectures lie in the fact that, in a session with N participants, each one sends and receives $N - 1$ audio streams. If the audio stream consists of uncompressed Pulse Code Modulated (PCM) stereo audio signals sampled at either 44.1 or 48 kHz, the required uplink/downlink bitrate is around 1.5 Mbps per participant. More sophisticated configurations, involving more than two channels per user, further increase such figures. It follows that the uplink bitrate offered to residential customers by Internet Service Providers (ISPs) may be insufficient for peer-to-peer NMP sessions involving a significant amount of participants. In such cases, a client-server configuration with a server node equipped with audio synchronization and mixing capabilities is typically leveraged. The server collects the audio streams generated by every musician in upstream, mixes them at the server's premises and produces a single stereo signal to be delivered in downstream, thus reducing bit-rate requirements per participant.

IV. TOWARDS AN INCLUSIVE FRAMEWORK FOR NMP AND REMOTE MUSICAL EDUCATION

Though a number of NMP solutions already exist, such as LOLA,[1] Jacktrip,[2] SoundJack,[3] JamKazam,[4] SonoBus,[5] FarPlay,[6] they show several limitations:

- Lack of accessibility for users with visual/auditory/mobility impairments.
- Lack of user-friendliness for subjects without technical background in communication technology (e.g., children): for instance, users may be requested to manually input IP addresses or UDP port numbers to enable multimedia streaming.
- Lack of portability and deployment flexibility: NMP solutions leverage wired networking infrastructures for the network access segment, due to the unacceptable latency introduced by the medium access control mechanisms of Wi-Fi and 4G technologies. The adoption of 5G connectivity for NMP practices has started being investigated very recently (see, e.g., [68]) but not yet fully integrated in NMP frameworks.
- Lack of a dedicated backend to support low-latency audio streaming, i.e., they treat latency contributions introduced by the telecommunication network as a ‘‘black box’, with no effort to minimize it to improve the perceived Quality of Service (QoS) and Quality of Experience (QoE).

To address the aforementioned shortcomings, our envisioned inclusive NMP platform for remote music teaching and practices (depicted in Fig. 1) adopts a client-server architectural paradigm. It envisions the possibility of leveraging a dedicated easy-to-use hardware (music box), with the goal of overcoming the real-time processing limitations of desktop computers and operating systems. The music box can leverage a single-board computer (e.g., a Raspberry-Pi, as in [69]), or even customizable hardware (e.g., a Field Programmable Gate Array board) to support the technologies described in the following.

A. INTERFACE CONTROL SYSTEM BASED ON A CAMERA FEEDBACK AND COMPUTER VISION, OR ON MOTION SENSORS

Enabling mobility-impaired users to provide a stream of control information to interact with the interfaces of the NMP platform and with remote performers requires wearable sensors and/or hand-tracking technologies. This represents a unique way for those users to participate in music performance and collaboration, overcoming barriers that prevent their involvement in traditional music-making settings.

User inputs can be captured by sensors tailored to the specific mobility abilities of the subject, such as:

- video-based hand pose reconstruction methods [70], or specialized unobtrusive sensors [71] to reconstruct small-scale hand and finger movements;
- specialized joysticks or buttons that can be actuated with minimal effort and range of motion using a hand or the chin;
- gaze tracking devices [72];
- sip-and-puff devices [73].

This would enable a mobility-impaired user to participate in music-making as a conductor or instructor, by providing a real-time stream of information to an ensemble of performers. The same input methodology can be used to allow the user to act as a performer [74], [75], [76], by controlling a synthetic instrument either by directly determining the notes being played, or by shaping some aspects of the sound of an instrument that operates, at least in part, automatically (e.g., a sequencer playing a predetermined melody or an arpeggiator).

Gestural data streams acquired by means of the previously described motion tracking techniques can be rendered at the remote counterparts either via realistic avatars or by using stylized, simplified representations that still convey the user's intention. For example, the avatar could reproduce the pose of a conductor/instructor whose movements match the inputs of the user. In addition to enhancing usability, this brings a significant advantage in terms of bit-rate requirements, since gestural data can be encoded much more efficiently in terms of data volume and with lower latency than video data [77], thus fostering scalability while meeting the strict latency requirements of NMP systems.

B. SENSORS AND ACTUATORS FOR REMOTE CONVEYANCE OF VIBRO-TACTILE SIGNALS

Haptic feedback mechanisms (possibly in combination with a touchscreen interface) can be adopted to ease perception

---

[1] https://lola.conts.it
[2] https://www.jacktrip.org
[3] https://www.soundjack.eu
[4] https://jamkazam.com
[5] https://sonobus.net
[6] https://farplay.io

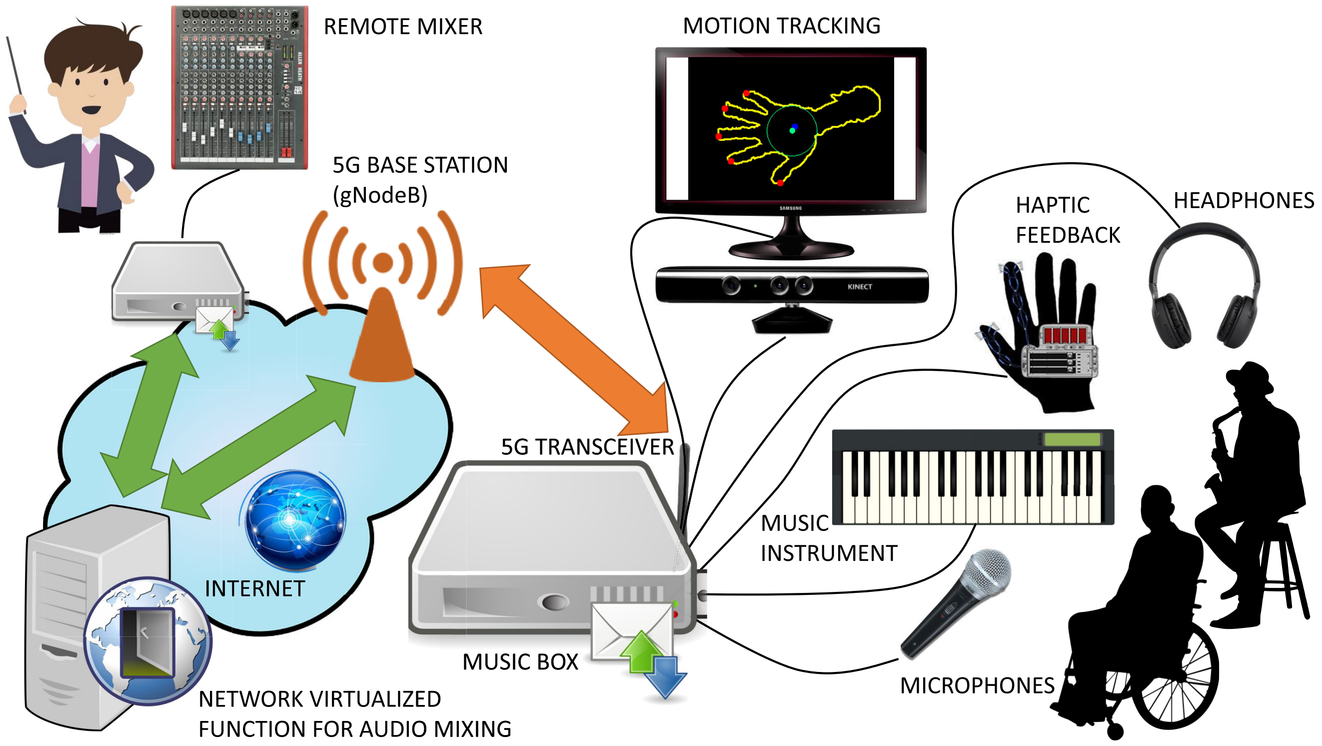

FIGURE 1. **The proposed inclusive NMP framework for remote music education and practice.**

for DHH and BVI users. Indeed, visually/auditory-impaired subjects heavily rely on haptic sensations to interact with the surrounding environment [78], [79]. Tactile wearable actuators [80] can also be used in combination with motion tracking systems (see previous paragraph) to convey haptic information regarding the gestures performed by the remote counterpart(s) to BVI/DHH users, e.g., those of a remote instructor or conductor.

C. ACCESSIBLE GRAPHICAL INTERFACE

The graphical user interface of the NMP system should be adherent to the accessibility standards for web applications [81] (e.g., compatibility with screen readers), possibly enhanced with tactile feedbacks [82] (e.g., by means of a touchscreen interface with customizable electroadhesive characteristics to render various textures in different areas of the screen, such as the device adopted in [83]).

D. REMOTE CONTROL OF MIXING FUNCTIONALITIES

Sound technicians should be enabled to remotely synchronize and mix the audio tracks received by all the participants to a NMP session, adjust their listening volumes and/or add audio effects such as equalization and reverb, to create personalized mixes for each performer. This would foster inclusion of subjects with reduced control capabilities of local interfaces due, e.g., to physical impairments.

E. VISORS FOR INTEGRATION OF AUGMENTED/VIRTUAL REALITY (AR/VR) ENVIRONMENTS WITH IMMERSIVE AUDIO RENDERING

The integration of spatialized audio rendering in the envisioned NMP framework is fundamental for conveying to the performance the sense of sharing a single acoustical scene [47]. Moreover, the usage of personalized HRTFs can help DHH users to compensate partial auditory deficiencies, with a moderate increase of bit-rate requirements. AR/VR environments can also be coupled to 3D audio rendering to further enhance the perception of being part of a shared musical environment [84].

F. 5G ACCESS NETWORK CONNECTIVITY

To interconnect the user's front-end to a dedicated backend infrastructure, 5G offers an unprecedented level of flexibility to fulfill service-specific requirements in terms of bitrate, latency, and reliability. 5G/beyond-5G wireless transmission technologies promise to guarantee access delay to the backbone telecommunication infrastructure below 10 ms, and even as low as 1 ms for mission-critical applications [85], which matches the latency requirements of NMP [86].

Additionally, flexibility of NMP deployments would be greatly enhanced by the adoption of 5G technologies. Indeed, the usage of cabled Ethernet connections for local area network connectivity heavily constrains portability, whereas Wi-Fi wireless connections are typically avoided in NMP setups, as they introduce excessively high jitter [87].

Finally, 5G networks are considered to be enablers of the so-called "tactile Internet" [85], which entails ultra-low latency transmission of haptic signals. Therefore, 5G/beyond-5G technologies constitute a promising candidate to accommodate the specific requirements of inclusive NMP systems. However, to date only a few designs of 5G infrastructures for NMPs have been investigated [88] along with a paucity of testbed deployments and in-depth statistical analysis on their latency and reliability performances [86]. Usage of 6G technologies is even more embryonal, though speculated by recent studies [89], [90].

At present, the features described in the above paragraphs have never been jointly designed and implemented in an NMP framework. Their potential usage and expected benefits are summarized in Table 2.

V. USE CASES

In this section, three illustrative use-cases for the framework presented in Section IV are discussed. All the inclusive features presented in the following are implemented in MEVO, an NMP system currently being developed in Politecnico di Torino and first demonstrated in June 2023 in a distributed concert involving musicians in Turin (Italy) and Wrocław (Poland) [91].

A. ACCESSIBLE GRAPHICAL INTERFACE

An accessible Graphical User Interface (GUI) is the first step towards an inclusive NMP system. Though several guidelines regarding how to implement accessible Web GUIs[7] are available, creating a fully accessible GUI requires considering several aspects. First of all, it should be easily navigable by a broad spectrum of users, including BVI and DHH subjects. To facilitate the navigation with assistive technologies such as screen readers, the GUI should be organized in a hierarchical structure that is logically coherent. This allows screen readers to parse and present the content in a well-defined structure, with the right level of importance to the various graphical elements, and allowing to skip recurring elements, such as navigation elements, to go directly to the main content of a page. Additionally, it must support keyboard-based interactions and provide meaningful descriptions for each element. Moreover, the GUI should be designed considering font sizes and colors contrast, to be distinguishable by subjects with color blindness or Color Vision Deficiency (CVD). Furthermore, it should adopt a terminology and implement/design interactions that ensure that any user is able to navigate and understand it, independently of his/her technological and educational background. All the above-mentioned aspects have been taken into account for the development of the GUI prototype of MEVO.

More specifically, MEVO's GUI allows the user to perform all the necessary actions to configure the system, create virtual rooms where users meet to play together, and control additional functionalities, such as remote mixing.

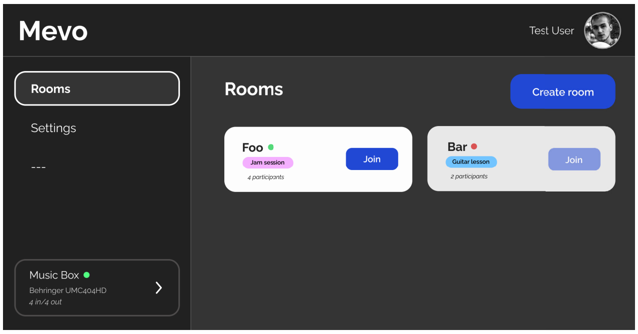

FIGURE 2. MEVO's accessible GUI - rooms page.

All the UI components on the web page are organized in a hierarchical structure, and each component was created following the web accessibility guidelines for UI components.[8] Following those guidelines ensures seamless integration of the GUI with assistive technologies, such as screen readers. We also made sure that color contrast was appropriate[9] for the various font sizes, to ensure readability by subjects with CVD.

The GUI presents a left sidebar, used for switching the various pages. The two main pages are: "Rooms" (to create virtual rooms) and "Settings" (to configure the various system settings). Furthermore, the room creation page allows for creating rooms with either selected participants or with a shareable link that anyone can use to join.

The creation of virtual rooms and the connection setup is handled and presented by the platform in a fashion similar to traditional videoconferencing applications. Moreover, any additional functionality to manipulate the audio stream, such as audio mixing, is presented with a layout that recalls the one adopted by traditional audio editing applications.

Once the users are connected to the same virtual room, each of them can individually customize the audio mixing of the various audio streams it receives from the other peers. If users are not able (or willing) to perform this action by themselves, it is possible to allow another participant (e.g., an assistant/audio engineer) to remotely control the audio mixing of each user. From a technical standpoint, this is possible thanks to the presence of the central server, which enables the communication between a user and a music box. Thus, the control messages a given box receives are independent of the user that generated them. This allows us to seamlessly handle switching from a local control to a remote control of a music box by another participant to the session, as long as he/she has the required permissions to do so.

B. ACCESSIBLE METRONOME

Metronomes can be easily integrated in NMP systems as done, e.g., in [92], even including adaptive capabilities to

---

[7] https://www.w3.org/WAI/standards-guidelines/wcag/

[8] https://www.w3.org/WAI/ARIA/apg/

[9] https://www.w3.org/WAI/WCAG21/Understanding/contrast-minimum

TABLE 2. Inclusive features of the proposed NMP framework and related benefits.

Feature	Usage	Expected Benefits
Accessible graphical interface enhanced with haptic feedbacks	Speech/gesture-enabled navigation of the system interface, haptic presentation of visual/audio content	Improve autonomous interaction with the NMP system of DVI/mobility-impaired users; complement the sensory perception of DHH users with audio visualization techniques enhanced with tactile feedbacks; offer haptic-enhanced score representation modalities to DVI users.
Motion tracking technologies	Acquisition of performers' gestures	Enabling of local control of virtual musical instruments (for motor-impaired and DVI users), conveyance of local gestures to remote counterparts and visualization via avatars without need of real-time video streaming.
Actuators for haptic signals	Remote conveyance of auditory or gestural cues	Support non-verbal interaction modalities with remote users, e.g., the tactile conveyance of the metronome beat (for DHH users) or of the conductor's gestures (for DVI users).
Remote control of mixing functionalities	Full operational integration of remote audio engineers/technicians	Availability of remote support in case of reduced control capabilities of local interfaces by disabled users (or by users without advanced technological background, e.g., children).
3D audio rendering and visors for AR/VR	Navigation of virtual spaces	Enabling 3D audio environment rendering and 3D visualization of virtual music instruments (for DHH and mobility-impaired users), possibility of customizing AR/VR spaces and the acoustic scene for educational and/or performative activities specifically tailored for the individual needs of the users.
5G connectivity	Support of low-latency, high-reliability wireless Internet access	Increased flexibility of deployment at the user's premises, overcoming physical barriers that may hinder accessibility to the system by disabled subjects.

time-varying latency conditions. In turn, audio cues generated by the metronome can be complemented with visual and tactile feedback to offer a multisensory experience that caters to a broad spectrum of users, including BVI and DHH subjects.

More in detail, our prototype of accessible metronome is designed to operate in a multi-user scenario where every user is provided with equipment consisting of a software component running on a PC, that interacts with a wearable device (a bracelet in our setting). Such system offers three distinct methods for delivering metronome cues:

- Graphical: a user-friendly GUI facilitates the interaction and provides a clear visualization of the metronome's beats in real-time for DHH users.
- Tactile: the bracelet is equipped with vibrating motors activated synchronously with the metronome beats for BVI or DHH users;
- Auditory: both the bracelet (through buzzers), and the PC (through speakers/headset) can emit sounds associated to metronome clicks.

Two different GUIs were designed. One acts as orchestrator and is meant to be run by only one user at a time to set up the settings shared among all the participants to the session. The available settings can be expressed in forms of single values (or lists of values), each of which describes a group of adjacent bars with common characteristics. The possible settings are:

- Number of bars: it indicates the amount of bars in each group (leaving this field empty configures the metronome to continue operating indefinitely).
- Bpm per bar: it indicates the beats per minute value for bars belonging to each group.
- Tempo signature per bar: it indicates the numerator of the tempo signature (i.e., the number of beats per bar) for each group.

Conversely, the second GUI displays two circles, each one mapping respectively a strong and a weak beat. The blue circle indicates the strong beats of a bar, while the red circle shows the weak beats. This visual representation aids users in identifying the timing of the metronome's beats and the bars subdivision.

The custom hardware component of the metronome is a bracelet which utilizes an Arduino Nano (with ESP32 in our setting) to control two vibrating motors and two buzzers, with two potentiometers to independently control their volume. This configuration was designed so that one motor and one buzzer are associated with strong beats of a measure, in accordance with the chosen time signature, while the second motor and buzzer were assigned to signal the remaining weak beats of the measure.

The software not only takes care of preserving the synchronization between the graphical beats and the bracelet through periodic feedback from the latter, but also exchanges real-time data with the participants to the online music session with the aim of preserving synchronization with all the participants' devices, minimizing the effects of network jitter and clock drifts.

C. TACTILE FEEDBACK FOR TIMBRE DISCRIMINATION

MEVO supports visual representation of pre-recorded audio samples enhanced with tactile feedback, with the aim of facilitating discrimination of timbre and sonic grain by DHH users. As thoroughly described in [83], audio excerpts generated by different instruments are associated with:

- a visual representation of the instrument;
- a background image and a background color, which should be evocative of the emotions aroused in the listener[10];
- a texture pattern superimposed on the background, to render the tactile sensation associated with the background image.

---

[10] The choice of colors and images to be associated with audio excerpts was taken based on the outcome of listening questionnaires, answered by 50 participants (see [83] for details).

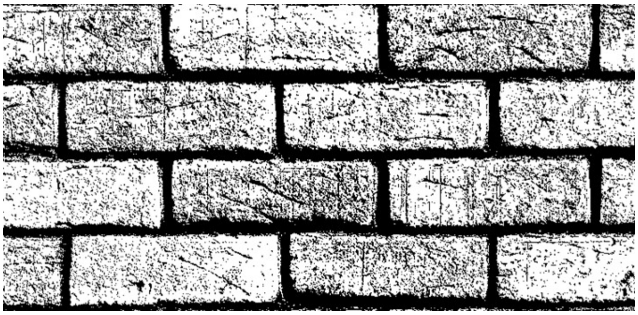

FIGURE 3. Example of black and white image rendering for haptic texture design of a brick wall [83].

FIGURE 4. Examples of visual representation of audio excerpts, enhanced with haptic feedback [83].

Cross-modal associations among images, colors, sounds, and textures are thus exploited to offer a multisensorial perceptual experience to the users.

The current prototype leverages a TanvasTouch device,[11] which enables a customized design of electroadhesion-based haptic feedback localized to the user's fingertip. Haptic textures are designed by means of black and white (B/W) images, where white corresponds to rough texture and black to smooth texture.

VI. TECHNICAL CHALLENGES

Based on the technology overview provided in Section II, it emerges that accessible web-mediated musical interactions in real-time are a rather unexplored area, with only a few studies investigating collaborative music making [93], mainly due to the extremely low latency requirements necessary to maintain a synchronous musical interplay [5]. Therefore, how to successfully incorporate such inclusive technological solutions within NMP frameworks for remote music teaching and practices is a yet almost completely unexplored research direction, which urges for dedicated efforts by the scientific community. A roadmap to guide such efforts is provided in the remainder of this Section, which reviews the most significant technical challenges involved in the realization of an inclusive NMP system, discussing open research directions.

A. LATENCY MANAGEMENT AND PACKET LOSS CONCEALMENT FOR CROSS-MODAL DATA STREAMS

When heterogeneous data types (e.g., audio, gestural and haptic signals) need to be jointly transmitted (i.e., they cannot be processed and streamed separately due to low-latency and high-reliability application requirements), cross-modal transmission strategies must be adopted [94].

After acquisition and digitization, data that need to be streamed over a network are packetized before being transmitted, to adhere to the packet-oriented switching paradigm adopted in telecommunication networks. Then, every packet is routed individually through the telecommunication network infrastructure to be conveyed to its intended destination. Consecutive packets may experience different delays along their end-to-end route, since some latency components (e.g., the queueing delays introduced by routers traversed by the packet along its path) may vary over time. Variations of the interarrival times between consecutive packets (the so-called "jitter") and packet losses due to transmission errors or drops (actuated by routers to prevent excessive traffic congestion) may cause data portions to arrive too late to be reproduced in due time at the receiver side. In turn, this generates glitches (in the case of audio streams), deterioration of the rendered gestures (for motion data) or of the perceived haptic feedback (for tactile data).

State-of-the-art packet retransmission techniques, such as those implemented by the TCP protocol at transport layer [95], cannot be exploited by NMP applications, as they would further increase the mouth-to-ear latency. Therefore, Packet Loss Concealment (PLC) techniques specifically tailored for cross-modal streams should be devised [96], [97], while ensuring compatibility with the low-latency requirements of NMP. Additionally, dedicated solutions must be developed to effectively handle cross-modal coding, transmission and latency when synchronizing audio, visual, and haptic data streams in real-time. A recent survey on such techniques can be found in [98].

Focusing on audio streaming, a further cause of artifacts is the so-called "drifting" effect caused by the imperfect alignment of local clock oscillators. Indeed, even when the clocks' nominal frequency is identical, small deviations may still occur due, e.g., to temperature variations or to fabrication imperfections. This leads to buffer over/underruns caused by the difference between the amount of audio samples generated by the sender and the amount of those reproduced by the receiver during a reference time window [99], [100]. Eventually, such differences cause either occasional drops of audio blocks or generate gaps in the audio stream. This side effect further motivates the need for adequate PLC mechanisms and/or for dedicated clock synchronization approaches (see e.g., [99], [100]).

B. INTEROPERABILITY AND STANDARDIZATION

The need for interoperability among the heterogeneous components that are envisioned to be part of an inclusive

---

[11] Tanvas. Surface Haptics. https://tanvas.co/technology

NMP system has already emerged in [101], which addresses Internet of Musical Things (IoMusT) environments. The lack of standardization of interfaces and communication protocols further hinders from seamless integration of Musical Things (i.e., networked devices capable of generating, playing, or responding to musical content) in NMP frameworks. Though some preliminary steps towards standardization for IoMusT have recently been done [102], the specific use-case of NMP has been neglected so far.

C. MINIMIZING NETWORK CONTRIBUTIONS TO MOUTH-TO-EAR LATENCY

The overall mouth-to-ear latency experienced by the users of NMP systems includes multiple contributions introduced by different stages of the audio acquisition, processing, transmission and reproduction chain. While acquisition, processing, buffering and playout delays can be controlled by NMP hardware/software designers, the latency component introduced by the network infrastructure is typically unknown and difficult to predict, as it is influenced by several factors such as the physical distance among performers, the network traffic congestion level (which exhibits variations over time), the traffic priority policies applied by network operators and ISPs, etc.

The exploitation of Software Defined Networking (SDN) [103], Network Function Virtualization (NFV) [104], and Multi-Access Edge Computing (MEC) [105] technologies is envisioned as a game-changer to ensure compliance to the QoS/QoE requirements of NMP systems. Such technologies should be incorporated in a backend infrastructure dedicated to NMP applications, which should include tailored functionalities capable of dynamically adapting the routing of audio streams and the positioning of NVFs hosting NMP servers by jointly considering:

- the physical location of the involved users and of the candidate sites for NVFs placement, possibly exploiting MEC to ensure user proximity;
- the current network congestion level;
- predictions on the future evolution of traffic patterns.

Moreover, the involvement of ISPs and network operators in the design of commercial NMP services would enable the adoption of prioritization criteria for NMP-generated data (e.g., with a dedicated SDN slice), thus offering adequate QoS guarantees to the users in terms of latency and jitter.

To implement such functionalities, the backend portion of the NMP system should also include a Service Management and Orchestration (SMO) engine, responsible for instantiating NVFs hosting NMP servers within MEC nodes and to coherently allocate resources in the involved network segments. This would enable latency optimization by locating NVFs in strategic positions within the network infrastructure, depending on the users' physical location, bitrate requirements, and on the traffic congestion level along the paths interconnecting users to servers. The SMO should also collect monitoring information from the users currently participating in the musical session (e.g., experienced latency and packet losses) to dynamically update the routing of audio flows as traffic conditions evolve. To this aim, the SMO could also exploit traffic prediction methods [106], [107] operating at various time scales. In turn, dedicated optimization algorithms implemented at the SMO premises will permit to reduce network latency by exploiting the outputs of such traffic predictors, which will enable proactive operations and dynamic adjustments of NVFs' locations depending on the current and forecasted network congestion conditions, thus ultimately improving the QoS/QoE perceived by the users.

Though a detailed discussion on such topics is out of the scope of this paper, the interested reader is referred to [45] for a proposal of an NMP infrastructure supporting immersive audio rendering, which leverages NFVs and MEC technologies. The adoption of SDN technologies for NMP systems has been first envisioned in [108], which provides examples of possible interactions between a real-time network latency monitoring module and a NMP system. Despite such preliminary attempts, the potential of SDN integration in NMP contexts remains largely unexpressed.

VII. FUTURE DIRECTIONS: TOWARDS AN INCLUSIVE MUSICAL METAVERSE

As of today, the concept of Musical Metaverse (MM), i.e., a virtual, digital world where each user musically interacts through his/her own avatar with other users, starts being postulated in the research community. According to [109, p. 5], the MM is "an interoperable persistent network of multiuser environments merging physical reality with digital reality, which serve a musical purpose. It is based on the convergence of Musical Extended Reality (Musical XR [18]) and Internet of Musical Things (IoMusT) technologies [110] that enable multisensory, networked musical interactions between musical stakeholders, as well as between such stakeholders and Musical XR environments and objects". The range of musical activities that could be performed in the MM encompasses composition, fruition of virtual performances, teaching, and collaborative content production.

In addition to IoMusT and Musical XR, the MM also builds upon a large pool of emerging technological advancements [109], including:

- *blockchain* and *Non-Fungible Tokens (NFTs)* to control, regulate and ensure transparency in copyright ownership of music-related material, as well as in royalties distribution;
- *digital twins* to create virtual musical entities replicating physical ones, such as music instruments, acoustic scenes etc.;
- *Artificial Intelligence and Machine Learning* to enhance user-experience by improving proactivity and context-awareness of musical services. To this aim, ML-based predictive algorithms leveraging user-gathered sensing data (e.g., musical gestures), can be exploited, as well as recognition/classification methods to categorize musical content.

- *5G-and-beyond communication infrastructure* to ensure seamless fruition of the MM by adhering to strict latency, bit-rate and reliability requirements while providing widespread access and supporting user mobility.

Among the opportunities that are envisioned to be unlocked by a widespread adoption of the MM, it is worth mentioning [109]:

- the growth of unprecedented forms of musical expressions and activities, entailing novel ways of engagement of the audience and fostering collaborative creation among performers;
- the constitution of new social communities and of novel music-based social interactions, with their own management and governmental rules;
- novel monetization opportunities for the music industry, related to music streaming and distribution in the MM.

It is also envisioned that inclusiveness will be one important characteristic of the MM, which will pave the way for novel opportunities for disabled musicians and audiences to access musical experiences [109]. For example, XR-based musical instruments could ease control and manipulation for visual/mobility-impaired users. Similarly, fruition and participation to music events in the MM could be extended to disabled subjects for whom physical participation to in-presence performances is hindered or precluded.

We believe that inclusive NMP systems will constitute a pivotal building block for the realization of a disability-friendly MM, by providing the technological substrate upon which such a virtual world will be built. Indeed, based on the MM framework proposed in [109], NMP systems will act as middleman to bridge the physical layer (constituted by users, Musical Things, etc.) and the link layer, which encompasses the networking and communication substrate, as well storage and computational nodes in the cloud. However, current NMP systems will need to be significantly evolved and adapted for seamless incorporation in MM ecosystems, thus calling for dedicated research efforts by the scientific community. In particular, 6G communication technologies combined with MEC functionalities are envisioned for future integration in NMP infrastructures, as they are capable of supporting real-time streaming of huge volumes of cross-modal data, thus finally enabling VR delivery for fully-immersive experiences [111].

VIII. CONCLUSION

This paper offered a thorough overview of the educational and performative needs of visually/auditory/mobility-impaired users in the context of remote music education and online musical practice. Technological solutions that could address such needs, which include spatialized audio rendering, music visualization, haptic feedback and motion tracking techniques were also discussed.

Moreover, the manuscript envisioned the integration of such technologies in networked music performance frameworks to enhance their inclusiveness, highlighting the technical challenges to be tackled and discussing future research directions towards the realization of an inclusive musical metaverse.

ACKNOWLEDGMENT

The authors would like to thank Prof. Luca Turchet (University of Trento), Prof. Chiara Nicora (Music Conservatory Guido Cantelli, Novara), Prof. Piera Bagnus (Music Conservatory Giuseppe Verdi, Como), and Prof. Alessandro Giusti (IDSIA-USI-SUPSI, Lugano) for the enlighting discussions. This manuscript reflects only the authors' views and opinions and the Italian Ministry of University and Research cannot be considered responsible for them. *(Matteo Sacchetto and Leonardo Severi contributed equally to this work.)*

REFERENCES

[1] A. Collins and R. Halverson, *Rethinking Education in the Age of Technology: The Digital Revolution and Schooling in America*. New York, NY, USA: Teachers College Press, 2018.
[2] A. D. Samala, S. Rawas, T. Wang, J. M. Reed, J. Kim, N.-J. Howard, and M. Ertz, "Unveiling the landscape of generative artificial intelligence in education: A comprehensive taxonomy of applications, challenges, and future prospects," *Educ. Inf. Technol.*, vol. 3, pp. 1–40, Aug. 2024.
[3] (2022). *National Center for Education Statistics, U.S. Department of Education*. [Online]. Available: https://nces.ed.gov/programs/coe/indicator/cgg
[4] I. Howell, K. J. Gautereaux, J. Glasner, N. Perna, C. Ballantyne, and T. Nestorova. (2020). *Preliminary Report: Comparing the Audio Quality of Classical Music Lessons Over Zoom, Microsoft Teams, VoiceLessonsApp, and Apple FaceTime*. [Online]. Available: https://www.ianhowellcountertenor.com/preliminary-report-testing-video-conferencing-platforms
[5] C. Rottondi, C. Chafe, C. Allocchio, and A. Sarti, "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, 2016, doi: 10.1109/ACCESS.2016.2628440.
[6] (2020). *European Music Schools in Times of Coronavirus*. [Online]. Available: http://www.musicschoolunion.eu/wp-content/uploads/2020/05/EMU-Survey-Coronavirus.pdf
[7] W. C. Payne, F. Ahmed, M. Zachor, M. Gardell, I. Huey, A. Hurst, and R. L. Dubois, "Empowering blind musicians to compose and notate music with soundcells," in *Proc. 24th Int. ACM SIGACCESS Conf. Comput. Accessibility*, Oct. 2022, pp. 1–14.
[8] J. Vetter, "Welle-a web-based music environment for the blind," in *Proc. Int. Conf. New Interface Musical Expression*, 2020, pp. 701–705.
[9] A. Saha, T. B. McHugh, and A. M. Piper, "Tutoria11y: Enhancing accessible interactive tutorial creation by blind audio producers," in *Proc. CHI Conf. Human Factors Comput. Syst.*, Apr. 2023, pp. 1–14.
[10] J. Abramo, "Disability in the classroom: Current trends and impacts on music education," *Music Educators J.*, vol. 99, no. 1, pp. 39–45, Sep. 2012.
[11] E. Frid and A. Ilsar, "Reimagining (accessible) digital musical instruments: A survey on electronic music-making tools," in *Proc. NIME*, 2021, pp. 1–26.
[12] E. Frid, "Accessible digital musical Instruments—A review of musical interfaces in inclusive music practice," *Multimodal Technol. Interact.*, vol. 3, no. 3, p. 57, Jul. 2019.
[13] P. Hash, "Teaching instrumental music to deaf and hard of hearing students," *Res. Issues Music Educ.*, vol. 1, pp. 1–24, Apr. 2003.
[14] F. Catania, P. Crovari, E. Beccaluva, G. De Luca, E. Colombo, N. Bombaci, and F. Garzotto, "Boris: A spoken conversational agent for music production for people with motor disabilities," in *Proc. 14th Biannual Conf. Italian SIGCHI Chapter*, Jul. 2021, pp. 1–5.
[15] J. A. Jellison and D. M. Taylor, "Attitudes toward inclusion and students with disabilities: A review of three decades of music research," *Bull. Council Res. Music Educ.*, vol. 2, pp. 9–23, Apr. 2007.
[16] S. K. Jones, "Teaching students with disabilities: A review of music education research as it relates to the individuals with disabilities education act," *Update: Appl. Res. Music Educ.*, vol. 34, no. 1, pp. 13–23, Nov. 2015.

[17] T. Laes and H. Westerlund, "Performing disability in music teacher education: Moving beyond inclusion through expanded professionalism," *Int. J. Music Educ.*, vol. 36, no. 1, pp. 34–46, Feb. 2018.

[18] L. Turchet, R. Hamilton, and A. Çamci, "Music in extended realities," *IEEE Access*, vol. 9, pp. 15810–15832, 2021.

[19] L. Turchet, C. Fischione, G. Essl, D. Keller, and M. Barthet, "Internet of Musical things: Vision and challenges," *IEEE Access*, vol. 6, pp. 61994–62017, 2018.

[20] F. Berthaut, "3D interaction techniques for musical expression," *J. New Music Res.*, vol. 49, no. 1, pp. 60–72, Jan. 2020.

[21] S. Serafin, A. Adjorlu, N. Nilsson, L. Thomsen, and R. Nordahl, "Considerations on the use of virtual and augmented reality technologies in music education," in *Proc. IEEE Virtual Reality Workshop K-12 Embodied Learn. through Virtual Augmented Reality (KELVAR)*, Mar. 2017, pp. 1–4.

[22] B. Remache-Vinueza, A. Trujillo-León, M. Zapata, F. Sarmiento-Ortiz, and F. Vidal-Verdu, "Audio-tactile rendering: A review on technology and methods to convey musical information through the sense of touch," *Sensors*, vol. 21, no. 19, p. 6575, Sep. 2021.

[23] A. Flores Ramones and M. S. del-Rio-Guerra, "Recent developments in haptic devices designed for hearing-impaired people: A literature review," *Sensors*, vol. 23, no. 6, p. 2968, Mar. 2023.

[24] L. Turchet and M. Barthet, "Co-design of musical haptic wearables for electronic music Performer's communication," *IEEE Trans. Hum.-Mach. Syst.*, vol. 49, no. 2, pp. 183–193, Apr. 2019.

[25] A. Haynes, J. Lawry, C. Kent, and J. Rossiter, "FeelMusic: Enriching our emotive experience of music through audio-tactile mappings," *Multimodal Technol. Interact.*, vol. 5, no. 6, p. 29, May 2021.

[26] D. Carvalho, J. Barroso, and T. Rocha, "Multisensory experience for people with hearing loss: A preliminary study using haptic interfaces to sense music," in *Proc. Int. Conf. Human-Computer Interact.*, 2022, pp. 292–306.

[27] T. Grosshauser and T. Hermann, "Augmented haptics—An interactive feedback system for musicians," in *Proc. Int. Conf. Haptic Audio Interact. Design*, 2009, pp. 100–108.

[28] M. D. Fletcher, "Can haptic stimulation enhance music perception in hearing-impaired listeners?" *Frontiers Neurosci.*, vol. 15, pp. 1–24, Aug. 2021.

[29] I. B. Gorbunova and A. A. Govorova, "Music computer technologies as a means of teaching the musical art for visually-impaired people," in *Int. Conf. Proc.*, 2018, pp. 19–22.

[30] D. Antón, G. Kurillo, A. Goñi, A. Illarramendi, and R. Bajcsy, "Real-time communication for Kinect-based telerehabilitation," *Future Gener. Comput. Syst.*, vol. 75, pp. 72–81, Oct. 2017.

[31] A. Hadjakos, "Pianist motion capture with the Kinect depth camera," in *Proc. Sound Music Comput. Conf.*, 2012, pp. 303–310.

[32] A. Hadjakos, T. Großhauser, and W. Goebl, "Motion analysis of music ensembles with the Kinect," in *Proc. Conf. New Interface Musical Expression*, 2013, pp. 106–110.

[33] M. Warburton, M. Mon-Williams, F. Mushtaq, and J. R. Morehead, "Measuring motion-to-photon latency for sensorimotor experiments with virtual reality systems," *Behav. Res. Methods*, vol. 55, no. 7, pp. 3658–3678, Oct. 2022.

[34] H. B. Lima, C. G. R. D. Santos, and B. S. Meiguins, "A survey of music visualization techniques," *ACM Comput. Surveys*, vol. 54, no. 7, pp. 1–29, Sep. 2022.

[35] J. Kim, S. Ananthanarayan, and T. Yeh, "Seen music: Ambient music data visualization for children with hearing impairments," in *Proc. 14th Int. Conf. Interact. Design Children*, Jun. 2015, pp. 426–429.

[36] J. A. Deja, A. Dela Torre, H. J. Lee, J. F. Ciriaco, and C. M. Eroles, "ViTune: A visualizer tool to allow the deaf and hard of hearing to see music with their eyes," in *Proc. Extended Abstr. CHI Conf. Human Factors Comput. Syst.*, Apr. 2020, pp. 1–8.

[37] T. B. McHugh, A. Saha, D. Bar-El, M. Worsley, and A. M. Piper, "Towards inclusive streaming: Building multimodal music experiences for the deaf and hard of hearing," in *Proc. Extended Abstr. CHI Conf. Human Factors Comput. Syst.*, May 2021, pp. 1–7.

[38] R. Pelzer, M. Dinakaran, F. Brinkmann, S. Lepa, P. Grosche, and S. Weinzierl, "Head-related transfer function recommendation based on perceptual similarities and anthropometric features," *J. Acoust. Soc. Amer.*, vol. 148, no. 6, pp. 3809–3817, Dec. 2020.

[39] C. Frauenberger and M. Noisternig, "3D audio interfaces for the blind," in *Proc. Workshop Nomadic Data Services Mobility*, 2003, pp. 11–12. [Online]. Available: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=02030c97aa1e989568a49b3af224f3c1e55de2b6

[40] B. F. G. Katz and L. Picinali, "Spatial audio applied to research with the blind," in *Advances in Sound Localization*, P. Strumillo, Ed., Rijeka, Croatia: InTech, Apr. 2011, pp. 225–250. [Online]. Available: https://www.researchgate.net/profile/Lorenzo-Picinali/publication/324963875_Spatial_Audio_Applied_to_Research_with_the_Blind/links/5bbb402092851c7fde33fd60/Spatial-Audio-Applied-to-Research-with-the-Blind.pdf

[41] A. Afonso-Jaco and B. F. G. Katz, "Spatial knowledge via auditory information for blind individuals: Spatial cognition studies and the use of audio-VR," *Sensors*, vol. 22, no. 13, p. 4794, Jun. 2022.

[42] J. R. Blum, M. Bouchard, and J. R. Cooperstock, "What's around me? spatialized audio augmented reality for blind users with a smartphone," in *Proc. 8th Int. ICST Conf.*, 2011, pp. 49–62.

[43] A. Arfaoui et al., "Designing interactive and immersive multimodal installations for people with disability," in *Virtual Reality and Its Application in Education*, D. Cvetkovic, Ed., Rijeka, Croatia: IntechOpen, Jan. 2021, doi: 10.5772/intechopen.80114.

[44] P. Cairns, H. Daffern, and G. Kearney, "Parametric evaluation of ensemble vocal performance using an immersive network music performance audio system," *J. Audio Eng. Soc.*, vol. 69, no. 12, pp. 924–933, Dec. 2021.

[45] 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 J. Commun. Soc.*, vol. 5, pp. 4691–4709, 2024.

[46] L. Turchet and M. Tomasetti, "Immersive networked music performance systems: Identifying latency factors," in *Immersive and 3D Audio: From Architecture To Automotive (I3DA)*. Piscataway, NJ, USA: IEEE Press, 2023, pp. 1–6.

[47] R. Hupke, S. Preihs, and J. Peissig, "Immersive room extension environment for networked music performance," *J. Audio Eng. Soc.*, no. 28, Oct. 2022. [Online]. Available: https://aes2.org/publications/elibrary-page/?id=21909

[48] M. Giordano, J. Sullivan, and M. M. Wanderley, "Design of vibrotactile feedback and stimulation for music performance," in *Series on Touch and Haptic Systems*. Springer, 2018, pp. 193–214.

[49] A.-A. Darrow, "Music and the hearing impaired: A review of the research with implications for music educators," *Update: Appl. Res. Music Educ.*, vol. 7, no. 2, pp. 10–12, Mar. 1989.

[50] M. Mandanici, R. Di Filippo, and S. Delle Monache, "The discovery of interactive spaces: Learning by design in high school music technology classes," *Technol., Knowl. Learn.*, vol. 26, no. 4, pp. 1131–1151, Dec. 2021.

[51] V. F. Martins, L. Gomez, and A. G. D. Correa, "Teaching children musical perception with MUSIC-AR," *EAI Endorsed Trans. e-Learning*, vol. 2, no. 5, p. 3, Mar. 2015.

[52] E. D. Innocenti, M. Geronazzo, D. Vescovi, R. Nordahl, S. Serafin, L. A. Ludovico, and F. Avanzini, "Mobile virtual reality for musical genre learning in primary education," *Comput. Educ.*, vol. 139, pp. 102–117, Oct. 2019.

[53] A. Barmpoutis, R. Faris, L. Garcia, L. Gruber, J. Li, F. Peralta, and M. Zhang, "Assessing the role of virtual reality with passive haptics in music conductor education: A pilot study," in *Proc. 12th Int. Conf.*, 2020, pp. 275–285.

[54] A. King, H. Prior, and C. Waddington-Jones, "Connect resound: Using online technology to deliver music education to remote communities," *J. Music, Technol. Educ.*, vol. 12, no. 2, pp. 201–217, Sep. 2019.

[55] P. M. Hash, "Remote learning in school bands during the COVID-19 shutdown," *J. Res. Music Educ.*, vol. 68, no. 4, pp. 381–397, Jan. 2021.

[56] D. Calderón-Garrido and J. Gustems-Carnicer, "Adaptations of music education in primary and secondary school due to COVID-19: The experience in Spain," *Music Educ. Res.*, vol. 23, no. 2, pp. 139–150, Mar. 2021.

[57] E. J. O'Leary and J. K. Bannerman, "Technology integration in music education and the COVID-19 pandemic," *Bull. Council Res. Music Educ.*, vol. 2, no. 238, pp. 23–40, Oct. 2023.

[58] N. C. Onuora-Oguno and D. S. Umeojiaka, "The dominance of e-learning and information communication technology (ict) in Nigerian music education: Problems and prospects," *J. Afr. Stud. Sustain. Develop.*, vol. 3, no. 1, pp. 1–26, 2020.

[59] P. Paul, P. S. Aithal, S. Sharma, and R. Saavedra, "Infrastructure-technology, socio-economical issues and challenges in ICT-based education and digital education with possible solutions—A scientific observation," in *Latest Concern and Research in Applied Social Science, Management in Digital & ICT Society*, P. K. Paul et al., Eds., New Delhi, India: New Delhi Publishers, 2023, pp. 1–16.

[60] M. S. Taat and A. Francis, "Factors influencing the students' acceptance of e-learning at teacher education institute: An exploratory study in Malaysia," *Int. J. Higher Educ.*, vol. 9, no. 1, pp. 133–141, 2020.

[61] M. G. Salimon, S. M. M. Sanuri, O. A. Aliyu, S. Perumal, and M. M. Yusr, "E-learning satisfaction and retention: A concurrent perspective of cognitive absorption, perceived social presence and technology acceptance model," *J. Syst. Inf. Technol.*, vol. 23, no. 1, pp. 109–129, Jun. 2021.

[62] D. Nie, E. Panfilova, V. Samusenkov, and A. Mikhaylov, "E-learning financing models in Russia for sustainable development," *Sustainability*, vol. 12, no. 11, p. 4412, May 2020.

[63] N. Hanes and S. Lundberg, "E-learning as a regional policy tool: Principles for a cost-benefit analysis," *RUSC. Universities Knowl. Soc. J.*, vol. 5, no. 1, p. 39, Apr. 2008.

[64] R. Setiawan, A. Putranto, Wihendro, E. Princes, I. Geraldina, E. Julianti, J. Safitri, and P. Pannen, "Tech-driven transformation: Innovative pricing strategies for E-Learning," *IEEE Access*, vol. 12, pp. 59063–59078, 2024.

[65] J.-P. Cáceres, R. Hamilton, D. Iyer, C. Chafe, and G. Wang, "To the edge with China: Explorations in network performance," in *Proc. 4th International Conference Digital Arts*, 2008, pp. 61–66.

[66] B. Redman, "The potential of videoconferencing and low-latency (LOLA) technology for instrumental music teaching," *Music Pract.*, vol. 6, p. 15, 2020.

[67] C. Drioli, C. Allocchio, and N. Buso, "Networked performances and natural interaction via LOLA: Low latency high quality A/V streaming system," in *Proc. Int. Conf. Inf. Technol. Perform. Arts, Media Access, Entertain.*, 2013, pp. 240–250.

[68] L. Turchet and P. Casari, "Latency and reliability analysis of a 5G-enabled Internet of Musical things system," *IEEE Internet Things J.*, vol. 2, no. 2, pp. 1–18, May 2023.

[69] C. Chafe and S. Oshiro, "Jacktrip on raspberry pi," in *Proc. 17th Linux Audio Conf. (LAC-19)*, 2019, pp. 12–22.

[70] W. Chen, C. Yu, C. Tu, Z. Lyu, J. Tang, S. Ou, Y. Fu, and Z. Xue, "A survey on hand pose estimation with wearable sensors and Computer-Vision-Based methods," *Sensors*, vol. 20, no. 4, p. 1074, Feb. 2020.

[71] O. Glauser, S. Wu, D. Panozzo, O. Hilliges, and O. Sorkine-Hornung, "Interactive hand pose estimation using a stretch-sensing soft glove," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–15, Aug. 2019.

[72] M. Cognolato, M. Atzori, and H. Muller, "Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances," *J. Rehabil. Assistive Technol. Eng.*, vol. 5, pp. 1–13, Jan. 2018.

[73] J. C. da Silva Junior and W. Germanovix, "Development of a sip-and-puff interface for communication and control of devices," in *Proc. Latin Amer. Conf. Biomed. Eng. Nat. Conf. Biomed. Eng.*, 2019, pp. 1137–1146.

[74] Q. Jarvis-Holland, C. Cortez, Nathan, Gamill, and F. Botello, "Expanding access to music technology - rapid prototyping accessible instrument solutions for musicians with intellectual disabilities," 2020, *arXiv:2011.09143*.

[75] A. Skuse and S. Knotts, "Creating an online ensemble for home based disabled musicians: Why disabled people must be at the heart of developing technology," in *Proc. Conf. New Interface Musical Expression*, 2020, pp. 1–12.

[76] S. Thorn, "Telematic wearable music: Remote ensembles and inclusive embodied education," in *Proc. Audio Mostly*, Sep. 2021, pp. 188–195.

[77] V. J. K. Morinigo, S. Benatti, and L. Benini, "A high SNR, low-latency dry EMG acquisition system for unobtrusive HMI devices," in *Proc. IEEE Biomed. Circuits Syst. Conf. (BioCAS)*, Oct. 2022, pp. 544–548.

[78] J. Kreimeier and T. Götzelmann, "Two decades of touchable and walkable virtual reality for blind and visually impaired people: A high-level taxonomy," *Multimodal Technol. Interact.*, vol. 4, no. 4, p. 79, Nov. 2020.

[79] M. D. Fletcher, "Using haptic stimulation to enhance auditory perception in hearing-impaired listeners," *Expert Rev. Med. Devices*, vol. 18, no. 1, pp. 63–74, Jan. 2021.

[80] Y. Chen, Y. Yang, M. Li, E. Chen, W. Mu, R. Fisher, and R. Yin, "Wearable actuators: An overview," *Textiles*, vol. 1, no. 2, pp. 283–321, Aug. 2021.

[81] R. G. Crespo, J. P. Espada, and D. Burgos, "Social4all: Definition of specific adaptations in Web applications to improve accessibility," *Comput. Standards Interface*, vol. 48, pp. 1–9, Nov. 2016.

[82] W. Safi, F. Maurel, J.-M. Routoure, P. Beust, M. Molina, C. Sann, and J. Guilbert, "Blind navigation of Web pages through vibro-tactile feedbacks," in *Proc. 25th ACM Symp. Virtual Reality Softw. Technol.*, Nov. 2019, pp. 1–12.

[83] M. Sacchetto, M. Sangüesa, P. Bagnus, C. Nicora, and C. Rottondi, "Collection of design directions for the realization of a visual interface with haptic feedback to convey the notion of sonic grain to DHH students," in *Proc. 4th Int. Symp. Internet Sounds*, Oct. 2023, pp. 1–7.

[84] B. Loveridge, "Networked music performance in virtual reality: Current perspectives," *J. Netw. Music Arts*, vol. 2, no. 1, pp. 2–12, 2020.

[85] S. S. Sefati and S. Halunga, "Ultra-reliability and low-latency communications on the Internet of Things based on5Gnetwork: Literature review, classification, and future research view," *Trans. Emerg. Telecommun. Technol.*, vol. 34, no. 6, Jun. 2023, Art. no. e4770.

[86] J. Dürre, N. Werner, S. Hämäläinen, O. Lindfors, J. Koistinen, M. Saarenmaa, and R. Hupke, "In-depth latency and reliability analysis of a networked music performance over public 5G infrastructure," in *Proc. Audio Eng. Soc. Conv.*, 2022, pp. 1–26.

[87] L. Turchet and E. Rinaldo, "Technical performance assessment of the ableton link protocol over Wi-Fi," *J. Audio Eng. Soc.*, vol. 69, no. 10, pp. 748–756, Oct. 2021.

[88] M. Centenaro, P. Casari, and L. Turchet, "Towards a 5G communication architecture for the Internet of musical things," in *Proc. 27th Conf. Open Innov. Assoc. (FRUCT)*, Sep. 2020, pp. 38–45.

[89] L. Mohjazi, B. Selim, M. Tatipamula, and M. A. Imran, "The journey toward 6G: A digital and societal revolution in the making," *IEEE Internet Things Mag.*, vol. 7, no. 2, pp. 119–128, Mar. 2024.

[90] X. Bai, "Remote music learning based on wireless sensors supporting 6G and CPS," *Wireless Pers. Commun.*, vol. 2, pp. 1–20, May 2024.

[91] L. Severi, M. Sacchetto, A. Bianco, C. Rottondi, A. Knapinska, and P. Lechowicz, "Demonstration of a networked music performance experience with MEVO," 2024, *arXiv:2404.09665*.

[92] R. Battello, L. Comanducci, F. Antonacci, G. Cospito, and A. Sarti, "Experimenting with adaptive metronomes in networked music performances!" *J. Audio Eng. Soc.*, vol. 69, no. 10, pp. 737–747, Oct. 2021.

[93] R. Hamilton and C. Platz, "Gesture-based collaborative virtual reality performance in carillon," in *Proc. Int. Comput. music Conf.*, 2016, pp. 337–340.

[94] X. Wei, M. Zhang, and L. Zhou, "Cross-modal transmission strategy," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 3991–4003, Jun. 2022.

[95] M. Allman, V. Paxson, and E. Blanton, *TCP Congestion Control*, document RFC 5681, 5681.

[96] X. Wei, Y. Yao, H. Wang, and L. Zhou, "Perception-aware cross-modal signal reconstruction: From audio-haptic to visual," *IEEE Trans. Multimedia*, vol. 2, pp. 1–12, 2022.

[97] Y. Chen, P. Li, A. Li, D. Wu, L. Zhou, and Y. Qian, "Toward low-latency cross-modal communication: A flexible prediction scheme," *IEEE Trans. Mobile Comput.*, vol. 23, no. 12, pp. 13310–13324, Dec. 2024.

[98] X. Wei, D. Wu, L. Zhou, and M. Guizani, "Cross-modal communication technology: A survey," *Fundam. Res.*, vol. 2, pp. 1–26, Sep. 2023.

[99] C. Werner and R. Kraneis, "UNISON: A novel system for ultra-low latency audio streaming over the Internet," in *Proc. IEEE 18th Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2021, pp. 1–4.

[100] P. Ferguson, C. Chafe, and S. Gapp, "Trans-Europe express audio: Testing 1000 mile low-latency uncompressed audio between Edinburgh and Berlin using gps-derived word clock, first with jacktrip then with dante," in *Audio Engineering Society Convention*, 2020.

[101] L. Turchet and F. Antoniazzi, "Semantic web of musical things: Achieving interoperability in the Internet of Musical things," *J. Web Semantics*, vol. 75, Jan. 2023, Art. no. 100758.

[102] R. Vieira and F. Schiavoni, "Sunflower: An environment for standardized communication of IoMusT," in *Proc. Audio Mostly*, Sep. 2021, pp. 175–181.

[103] K. Nisar, E. R. Jimson, M. H. A. Hijazi, I. Welch, R. Hassan, A. H. M. Aman, A. H. Sodhro, S. Pirbhulal, and S. Khan, "A survey on the architecture, application, and security of software defined networking: Challenges and open issues," *Internet Things*, vol. 12, Dec. 2020, Art. no. 100289.

[104] B. Han, V. Gopalakrishnan, L. Ji, and S. Lee, "Network function virtualization: Challenges and opportunities for innovations," *IEEE Commun. Mag.*, vol. 53, no. 2, pp. 90–97, Feb. 2015.

[105] A. Filali, A. Abouaomar, S. Cherkaoui, A. Kobbane, and M. Guizani, "Multi-access edge computing: A survey," *IEEE Access*, vol. 8, pp. 197017–197046, 2020.

[106] K. Lee, M. Eo, E. Jung, Y. Yoon, and W. Rhee, "Short-term traffic prediction with deep neural networks: A survey," *IEEE Access*, vol. 9, pp. 54739–54756, 2021.

[107] D. A. Tedjopurnomo, Z. Bao, B. Zheng, F. M. Choudhury, and A. K. Qin, "A survey on modern deep neural network for traffic prediction: Trends, methods and challenges," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 4, pp. 1544–1561, Apr. 2022.

[108] E. Lakiotakis, C. Liaskos, and X. Dimitropoulos, "Improving networked music performance systems using application-network collaboration," *Concurrency Computation, Pract. Exper.*, vol. 31, no. 24, Dec. 2019, Art. no. e4730.

[109] L. Turchet, "Musical metaverse: Vision, opportunities, and challenges," *Pers. Ubiquitous Comput.*, vol. 27, no. 5, pp. 1811–1827, Oct. 2023.

[110] L. Turchet, F. Antoniazzi, F. Viola, F. Giunchiglia, and G. Fazekas, "The Internet of musical things ontology," *J. Web Semantics*, vol. 60, Jan. 2020, Art. no. 100548.

[111] J. Yu, A. Alhilal, P. Hui, and D. H. K. Tsang, "6G mobile-edge empowered metaverse: Requirements, technologies, challenges and research directions," 2022, *arXiv:2211.04854*.

CRISTINA ROTTONDI (Senior Member, IEEE) received the Ph.D. degree in information engineering from Politecnico di Milano, Italy, in 2014. From 2015 to 2018, she had a research appointment with the Dalle Molle Institute for Artificial Intelligence (IDSIA), Lugano, Switzerland. She is currently an Associate Professor with the Department of Electronics and Telecommunications, Politecnico di Torino, Italy. She is the co-author of more than 100 scientific publications in international journals and conferences. Her main research interests include optical network planning and networked music performance.

MATTEO SACCHETTO (Graduate Student Member, IEEE) received the master's degree (cum laude) in computer engineering from Politecnico di Torino, Italy, in 2020, where he is currently pursuing the Ph.D. degree with the Department of Electronics and Telecommunications. His master's thesis focuses on "JackTrip-WebRTC-Networked Music Performance with Web Technologies." His main research interests include networked music performance and WebRTC and web audio.

LEONARDO SEVERI (Graduate Student Member, IEEE) received the B.Sc. and M.Sc. degrees in computer engineering from Politecnico di Torino, Italy, in 2018 and 2021, respectively, and the Dipl.-Ing. degree from Grenoble INP-Ensimag, in 2021, where he is currently pursuing the Ph.D. degree with the Department of Electronics and Telecommunications. His current research interest includes networked music performance.

ANDREA BIANCO (Senior Member, IEEE) received the Ph.D. degree in telecommunications from Politecnico di Torino, Italy, in 1993. He is currently a Full Professor and the Vice Rector of Internal Affairs of Politecnico di Torino. He has co-authored more than 230 papers published in international journals and presented at leading international conferences in the area of telecommunication networks. His main research interests include the design of all-optical networks, switch architectures for high-speed networks and data centers, SDN, and networked music performance.

• • •

Open Access funding provided by 'Politecnico di Torino' within the CRUI CARE Agreement