Inclusiveness in Remote Music Teaching and Networked Music Performances: Vision, Technological Enablers and Design Strategies

(Article begins on next page)

Inclusiveness in Remote Music Teaching and Networked Music Performances: Vision, Technological Enablers and Design Strategies

Cristina Rottondi, *Senior Member, IEEE*

***Abstract*—Ensuring inclusiveness in networked musical education and practices raises unprecedented technical, artistic and pedagogical challenges. Due to its immense social and cultural impact, the research community is urged to tackle such challenges with holistic and multidisciplinary approaches, encompassing the know-how and competences of experts in several domains such as Information and Communication Technologies (ICT), pedagogy, psychology, and healthcare. Moreover, given the vast range of disabilities, the need for personalized solutions, designed for the specific needs and educational goals of each diversely able subject, clearly emerges.**

This paper reviews the main technological enablers for the design of tailored solutions to support the daily activities of music students and practitioners with visual, auditory or mobility impairments. Then, it discusses an interdisciplinary design methodology to devise technology-assisted educational activities adapted to the special needs of a given individual. Finally, the manuscript proposes some use-cases to exemplify the potential adoption of such solutions.

***Index Terms*— Inclusive education, remote music teaching, networked music performances.**

I. INTRODUCTION

THE way pedagogical processes are being conceived is undergoing dramatic changes due to the progressive adoption of Information and Communication Technologies (ICT). The last ten years have seen a proliferation in distance education and computer-based learning, and this trend is expected to continue in the forthcoming decade. This suggests the need for a structural reconsideration of the ways in which educational services are designed and operated.

Long-term adoption of technologies for remote learning would be advantageous for various categories of students. Among those, a non-negligible portion consists of subjects with special needs, e.g., due to impairments in vision, hearing, or mobility, for whom traditional in-person access to education is difficult or even impossible. Statics reported in [1] state that 7.2 million US students, accounting for roughly 15% of the entire public-school population, received special education services in 2020–2021. According to the same source, 1% of students with special needs were either homebound or confined in hospitals and separate residential facilities due to their health conditions. It follows that guaranteeing inclusiveness and equitable access to education has become one of the most pressing and impactful priorities of our century, as testified by the numerous legislative initiatives undertaken worldwide (see, e.g., the United Nations Convention on the Rights of Persons with Disabilities).

In this context, music education makes no exception: the challenge of inclusion in music-related pedagogical practices is currently a hot topic, which has generated a heated debate on what it actually entails. Notably, in comparison to other subjects of study, music teaching and learning involve different interaction modalities, which can hardly be accommodated by the typical technological supports adopted in remote educational settings. More in detail, remote music education may imply specific technical requirements, such as:

- strict Mouth-to-Ear (M2E) latency guarantees, in the range of a few tens of ms [2], to guarantee realistic performative conditions for music ensembles with geographically dispersed members;
- high data rate and transmission reliability, to ensure professional audio quality (i.e., streaming of uncompressed audio channels);
- adequate computational power for the processing, mixing and reproduction of multiple audio streams, to support large-scale deployments (e.g. for music classes, ensembles, orchestras and choirs).

Currently, two technological solutions can be leveraged for distance education in music-related areas: videoconferencing or Networked Music Performance (NMP) frameworks.

The former ones are primarily designed to maximize speech intelligibility. To this aim, low sampling frequencies and highly efficient compression codecs are adopted to reduce the necessary bit rate at the expense of the quality of voice signals, which may be subject to degradation. Moreover, due to their data compression algorithms, built-in audio codecs increase the perceived end-to-end delay due to the encoding/decoding latency. In contrast, resilience to packet losses and delays is increased thanks to the usage of large playout buffers, which introduce additional latency. Therefore, off-the-shelf videoconferencing systems are a viable solution for one-to-one teaching scenarios where in-sync playing is not required and compromises on the achievable audio quality are acceptable. Conversely, they are not directly applicable in contexts where

C.Rottondi is with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy. (e-mail: cristina.rottondi@polito.it).

ensemble playing or high-fidelity audio quality is essential.

Such shortcomings are instead addressed by NMP systems, which have flourished in the last decade, especially during the SarS-CoV-2 pandemic period. NMP frameworks enable physically distant musicians to interact in real time by taking advantage of low-latency uncompressed audio streaming over Wide Area Networks (WANs) [2]. However, despite successfully tackling the issue of low-latency audio streaming and ensuring professional audio quality, existing NMP frameworks are mainly targeted to performative activities, rather than on educational ones.

Unfortunately, the vast majority of videoconferencing platforms for remote education and of NMP systems are not designed following the paradigms of accessibility, thus resulting in reduced usability by subjects with special needs. Since such lack of accessibility prevents a diffused adoption of music-related remote educational practices and hinder their inclusiveness, this paper:

- offers an overview of features and technological enablers that could be leveraged by inclusive systems for distance education (surveyed in Section II);
- discusses the intertwining of low-latency multimedia streaming with the field of the Internet of Sounds and the emerging concept of Musical Metaverse (outlined section III);
- identifies a holistic and multidisciplinary design approach for the realization of personalized technological solutions, tailored to the special needs of users with disabilities (detailed in Section IV);
- showcases exemplary scenarios where the above-mentioned solutions are leveraged to support remote musical practices (described in Section V).

II. Features and Technological Enablers for Inclusive Musical Education and Practices

This Section provides a summary of technical features that inclusive systems for remote music learning could incorporate to enable a diversely able subject to musically interact with others, in ways that would otherwise be precluded. Though attempts to include such features in NMP frameworks have already been documented, their joint integration in a an “inclusive-by-design” NMP system remains to be investigated.

A. Haptic Feedback

Several studies investigated how the sense of touch can carry musical information (for a comprehensive review, see [3]). Typically, tactile rendering systems focus on capturing a single feature, such as loudness, rhythm, pitch, melody, or timbre.

The usage of musical haptic devices has been proposed to facilitate communication between artists sharing the same physical space, or to enhance the emotional experience aroused by music listening. Moreover, vibrotactile feedback has been shown to help Deaf and Hard of Hearing (DHH) and Blind and Visually Impaired (BVI) students interiorize the concept of rhythm and explore the spectrum of sound frequencies [4]. In the context of NMP systems, vibro-tactile feedback could be incorporated (in conjunction with motion tracking techniques, introduced in the next subsection) to remotely convey gestural cues or enhance the perception of musical timbre or of different frequency components of the conveyed audio signals for DHH and BVI users.

B. Motion Tracking

Numerous publications have already demonstrated the positive impact of the adoption of augmented and virtual reality for music education of students with disabilities [5]. Additionally, motion tracking technologies for sound and visual production have proved to facilitate musical expression of subjects with mobility or cognitive impairments [6].

From a networking perspective, streaming pose data (to be rendered via avatars) instead of video frames enables substantial reductions in transmission bit-rate requirements, since it only requires the conveyance of the Cartesian position and rotation data of the joints being represented. Therefore, savings are of at least one order of magnitude with respect to encoded video streaming and may reach three orders of magnitude when considering unencoded video. Instead, for what concerns the motion-to-photon latency (i.e., the time taken for a user gesture to result in a corresponding visual change on a display), experimental values are below 40 ms [7], i.e., comparable to acquisition and encoding/decoding delays introduced by a camera streaming compressed video frames.

As already mentioned, the pairing of motion tracking and haptic feedback technologies can be exploited in NMP frameworks for the remote conveyance of gestures (e.g., the ones of the music instructor, of an orchestra conductor or of the leading member of a music ensemble).

C. Music Visualization

Visualization techniques can be leveraged as alternative conveyance form to present music content to subjects with auditory impairments. A few studies have specifically focused on the DHH community, within the extensive body of scientific literature on this topic (see [8] for a thorough survey).

On-screen visualization is usually exploited to create real-time representations, alongside the reproduction of audio content. For example, sounds produced by traditional instruments can be digitally processed and linked to the movements of specific objects, in an effort to create an interactive sound visualization environment that may support music learning activities for DHH subjects. In NMP systems, visualization techniques can be exploited to enhance the perception by DHH users of specific features of the audio signals (e.g., pitch or timbre), but also to trigger emotions associated with the musical piece being played.

D. Immersive Audio Rendering

The goal of immersive audio systems is to provide the listener with a three-dimensional (3D) representation of the acoustic sources, so they can be perceived as emanating from a particular spatial direction. Headphones are more commonly used in the specific use-case of remote education, rather than surround sound systems, due to their lower cost and ease of portability and configuration. Binaural audio rendering is the immersive audio solution that is usually adopted in such a scenario.

For BVI users, spatial audio technology has already shown to improve the perception of virtual imagery and spatialization [9]. When it comes to musical practices, 3D audio rendering can be used to replicate a common acoustical environment at each remote participant's location. Moreover, by exploiting a personalized placement of spatialized sound sources, BVI subjects can improve their ability to discriminate among the auditory cues needed to synchronize their playing with that of the other performers. Similarly, DHH users might profit from an auditory scene rendered at their location that is customized for their specific hearing impairment. Thus, the integration of immersive audio rendering in NMP systems is deemed fundamental to ease audio perception by BVI and DHH subjects.

III. Low-latency Multimedia Streaming Technologies for Remote Musical Education

A. Networked Music Performance Systems

The fundamental function of NMP systems is enabling professional-quality real-time audio streaming, with M2E latency figures similar to those attained by musicians performing in the same physical location. Ideally, latency should not exceed 20–30 ms, corresponding to the time taken by a sound wave to cover 8–10 m. Beyond such threshold, the musicians' delayed perception of the sounds produced by their remote counterpart(s) starts negatively affecting musical interplay, typically leading to tempo deceleration [3]. NMP systems can operate either in peer-to-peer or client/server fashion: in comparison to the former, the latter ensures greater scalability for large deployments (i.e., exceeding a few participants), at the price of a slight latency increase due to the presence of the server operating as audio mixing and relay node. Other features that are commonly integrated in NMP systems include communication and interaction interfaces (chats, user status tracking, etc.), collaborative music composition, editing and production tools and local audio manipulation capabilities (e.g., recording, reproduction of pre-recorded streams, adjustment of volumes, equalization levels, and panning).

B. Videostreaming for Networked Music Performances

Video streaming features are included in some NMP solutions. However, typical video acquisition and encoding/decoding times greatly surpass the 30 ms latency threshold. As a result, during real-time remote musical interactions the provided video stream is typically out of sync with the corresponding audio stream, thus impeding reliance on visual cues. To bypass this issue, video can be streamed unencoded, which requires a communication infrastructure able to support uplink/downlink bitrates of a few Gbps. Since such figures are hardly reachable by residential Internet connectivity solutions, the widespread use of low latency unencoded video streaming is hindered.
Alternative solutions to drastically reduce bit-rate requirements are those based on avatar-based representations of the NMP participants in a virtual scenario (discussed in Section II.B), which only require the conveyance of gestural data and metadata related to the reconstruction of the virtual environment, instead of video frames.

C. Cross-modal Streaming for Internet of Musical Things Technologies

The Internet of Musical Things (IoMusT) 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 network [10]. It follows that IoMusT technologies can potentially be integrated in NMP frameworks. When heterogeneous intertwined streams such as audio, video, gestural, and tactile data gathered from Musical Things need to be transmitted in the network, they are packetized after being digitized. As packets travel through the network, they can face varying delays and losses, leading to issues like audio glitches, poor gesture rendering, or weak haptic feedback when the streamed content is reproduced. Traditional retransmission methods like those implemented by the Transmission Control Protocol (TCP) cannot be leveraged in such a scenario because they would excessively increase the M2E latency. Therefore, cross-modal transmission methods must be adopted, leveraging Packet Loss Concealment (PLC) techniques specifically designed for cross-modal data. Though this research area is still in its infancy, notable attempts to develop PLC models for low-latency audio and video reconstruction already exist. Among those, several ones leverage generative models to produce audio/video content to replace gaps in the received datastreams.

D. The Musical Metaverse as Novel Educational Landscape

Currently, the research community is exploring the idea of a Musical Metaverse (MM), envisioned as a digital universe where users can interact musically via avatars. As described in [11], the MM is an interconnection of shared spaces that blend the physical and digital worlds for musical activities. It relies on the fusion of Musical Extended Reality (Musical XR) and IoMusT technologies, which facilitate immersive, networked musical exchanges among participants and between them and the Musical XR settings and items. The MM could host a variety of musical endeavors, including composing, enjoying virtual concerts, teaching, and joint content creation.
The MM is anticipated to be a space that promotes inclusivity, creating new possibilities for musicians and audiences with disabilities to engage with music. For instance, musical instruments designed with XR could be more manageable for those with visual or mobility impairments. Similarly, the MM could make it easier for individuals with disabilities, who might find it challenging to attend live performances, to participate in musical events. Inclusive NMP systems are expected to be crucial for developing an MM that is accessible, serving as the foundational technology for this virtual environment.

IV. A Holistic Design Process for Technology-Assisted and Inclusive Music Educational Activities

Creating inclusive educational tools is challenging due to traditional frameworks and methodologies not fully addressing the needs of users with disabilities. Even when generic accessibility guidelines are incorporated, they often fall short in supporting specific physical or cognitive impairments. Moreover, in technology-assisted education, the potential role of pedagogists in the design workflow accessible e-learning tools is often overlooked. Incorporating healthcare professionals' knowledge is even more arduous, due to privacy and ethical barriers.

When focusing on the use-case of remote music education, ensuring inclusiveness becomes particularly difficult, due to the inherently multi-sensory nature and the incredibly nuanced variety of musical activities. Thus, developing accessible solutions customized to music-related e-learning activities is as a complex process that requires dedicated and specialized workflows.

In this paper, the DesignABILITY framework [12] is taken as reference for the design and development of hardware or software-based tools for users with special needs. DesignABILITY was conceived starting from the assumption that a given educational goal may be achieved with different modalities and learning strategies, depending on the particular requisites posed by the user condition. Therefore, the notion that learning tools and environments should be adaptively and flexibility adjusted to meet the needs of learners becomes pivotal.

The DesignABILITY framework envisions a workflow that consists of four phases: (1) definition of learning/performance requirements; (2) design of engaged learning/performance; (3) prototyping; (4) evaluation. In the following, each phase is detailed and customized to envision a holistic approach to the design process of inclusive educational tools for remote music teaching and practices. It must be noted that while the description of such phases is purposedly kept general, a proper tailoring of the workflow should be applied when dealing with single use-cases, possibly enriching it with multiple consecutive Design-Develop-Evaluate iterations to progressively increase the technology readiness level of the designed tools.

A. Definition of Learning/Performance Requirements

This stage aims at identifying the learning/performance goals to be supported and the teaching approaches to be adopted by the instructor. Goals and approaches will vary, depending on the characteristics of the diversity being tackled and on the specific needs of every particular user of the tool. In addition to instructors and pedagogists, this identification process may require the expertise of medical professionals (e.g., psychologists, psychiatrists, neurologists, physiotherapists). For example, the goal of involving a motor-impaired user in a musical performance by interacting with a virtual instrument may be declined in many different ways, depending on his/her motor capabilities. The musical activity may even be instrumental to a physical training/rehabilitation scope: in such case, medical specialists may assist pedagogists in devising strategies to couple educational and therapeutical objectives.

B. Design of Engaged Learning/Performance

After setting the learning targets, the way in which users will interact with the tool and be involved in the learning/performance activity must be defined. This step implies a personalized design, according to the special needs of the final user, and requires a joint effort by technicians, pedagogists and medical staff. For example, sound conveyance to a hard-of-hearing music student can be supported via tactile or visual feedback, or even with a combination of the two. In addition, binaural audio could be exploited to create a customized placement of audio sources in the rendered sonic landscape, tailored to the specific auditory impairment of the student and incorporating indications by his/her otolaryngologist.

Differently, an autistic child may benefit of the adoption of a storytelling that involves one of his/her special interests, since a typical trait of autistic subjects is the intense focus that individuals develop in topics they are strongly passionate about. In such a case, a visual interface that incorporates images pertaining such special interests may result more engaging.

Learning/performance targets and strategies should be translated in a set of technical specifications and requirements for the prototype to be developed in the next phase.

C. Prototyping

Once learning/performance goals and technical requirements have been collected, the realization of the tool can finally start. At this stage, the main challenge is how to effectively integrate a variety of technological enablers that have already been widely adopted in isolation, but never combined in inclusive frameworks for remote musical practices. In the prototyping phase, special attention should be given to ensure accessibility of the Grafical User Interface (GUI) and, more in general, to guarantee adequate quality of the User eXperience (UX) in terms of usefulness, ease of use, and desirability.

D. Evaluation

The last phase of the process entails the assessment of the designed tool, which encompasses experts' reviewing and usability testing. Concerning the former, the same categories of experts that were involved in phase A (i.e., pedagogists and medical professionals) are called to evaluate the appropriateness of the tool for the fulfilment of the educational goals that were initially set and the compliance of the tool to specific usability requirements identified during phase B.

Moreover, a variety of usability tests should be executed, e.g. by observing the users while interacting with the tool, administering questionnaires (e.g., the User Experience Questionnaire – UEQ, questionnaires to measure the System Usability Scale - SUS, or custom-made ones), and collecting opinions by means of ratings in Likert-scale.

V. Exemplary Showcases

In this section, the four design phases described in Section IV are declined in three envisioned scenarios involving remote musical practices. For each showcase, educational goals, technical and interaction requirements, prototype components, and assessment mechanisms are summarized in Table I. Educational goals were defined by means of interviews with three experts in inclusive music teaching.

A. First Scenario

Alice is a blind violin student attending the regular study program offered by a local music Conservatory. Though she attends most of her lessons in presence, she needs to practice regularly with her string quartet, composed by other students of her class, out of the regular lessons schedule. Since Alice cannot travel independently, participating to rehearsal sessions is logistically challenging for her, thus she seeks for a way to rehearse from home.

To this aim, she leverages the setup depicted in Figure 1, consisting of an NMP system involving two locations: Alice's room and the Conservatory classroom where the other three members of the quartet usually gather. The NMP system may operate either in peer-to-peer or client/server fashion and should implement adequate PLC techniques to guarantee high-fidelity audio streaming. While rehearsing, Alice wears headphones for binaural audio rendering, which reproduces the exact location of sound sources that she would perceive when playing in-presence with her classmates, in their regular setup, thus reconstructing a familiar acoustic environment to strengthen her self-confidence and help her in the fine-tuning of her instrument. The placement of the three sound sources corresponding to the three remote quartet members in the 3D audio landscape should be done beforehand, during an in-presence session, by the music instructor or a quartet member.

The quartet also makes use of a metronome reference during rehearsals, which can be easily integrated in the NMP system as done in [13]. In addition, to synchronize the start of the performance and to share indications about agogic and dynamics of the piece in absence of visual cues, Alice and the remote first violin wear a bracelet equipped with motion sensors and vibrational motors. One of the bracelets acquires motion data, which are collected by the NMP system (e.g. via a wired connection to the device where the NMP software is executed) and then streamed to the device worn by remote counterpart via the WAN. The latter bracelet is configured to produce vibrations when a movement occurs. This way, the first violin can remotely convey gestural cues as guidance for Alice.

Finally, the GUI of the NMP system must be fully compatible with screen readers, to guarantee full accessibility to BVI users. Moreover, to facilitate Alice in the correct setup of audio settings, remote control of volumes, mixing, panning and reverberation are implemented as added functionalities to the NMP system, so that her instructor or one of the quartet members are enabled to adjust audio levels if necessary.

The effectiveness of the above-described NMP system can be assessed by the members of the quartet and their instructor by means of questionnaires covering the following key-aspects: (1) perceived audio quality; (2) impact of audio streaming delay; (3) ease of use; (4) immersiveness of the NMP experience; (5) level of engagement; (6) quality of the rehearsing experience.

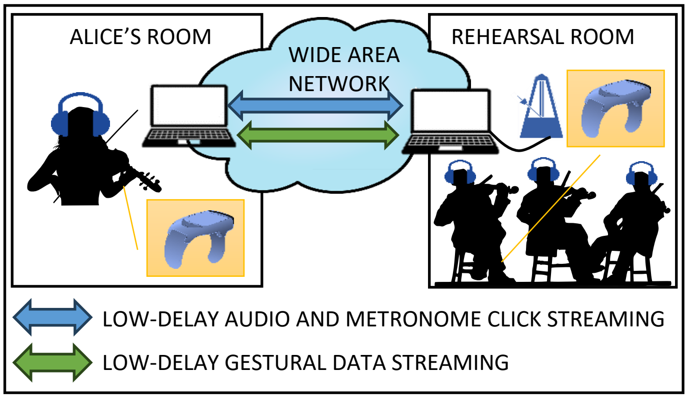


Fig. 1. NMP setup for remote rehearsals with a blind user.

B. Second Scenario

Bob is a first-grade school pupil with a mild, unilateral hearing loss in his left ear, partially compensated by wearing a hearing aid. The music instructor of his class wants to teach how to recognize music instruments based on their timbral characteristics. To do so, he adopts as tool a visual interface enhanced with haptic feedback, such as the one described in [14]. The tactile feedback perceived by the user when moving the fingertip across the screen surface can be customized by superimposing a black and white pattern to the image visualized on the screen itself. According to such pattern, the electroadhesive properties of the screen surface are set, reproducing arbitrarily designed textures, ranging from flat (black color) to rough (white color). A dedicated software application is then programmed to offer a tactile feedback-enhanced visual representation of pre-recorded audio samples. More in detail, each audio excerpt is associated with a background color, an image and a texture, as reported in Fig. 2. Bob's instructor can select such combinations beforehand, among a set of pre-loaded options. By exploiting such cross-modal associations between colors, textures, sounds, and images, the instructor can more effectively convey the notion of musical timbre to Bob and his classmates by letting them navigate the application while enjoying a multisensory experience.

Moreover, if the visual interface is remotely controllable (e.g., because it is integrated in an NMP system), Bob's instructor could more easily be involved in the auditory rehabilitation sessions that Bob attends on weekly basis, thus being able to reconfigure the interface in team-work with Bob's therapist and even to provide additional audio stimuli generated in real-time, in addition to the pre-loaded ones (e.g., playing notes in a specific range of frequencies with a given instrument), without need to physically attend the session. The streaming of uncompressed audio enabled by the NMP system is particularly important if exposition to high frequencies (i.e., above 15/20 KHz) is required, as that range is typically most impacted by compression techniques used by off-the-shelf videoconferencing technologies.

Furthermore, binaural audio rendering can be tailored to Bob's hearing loss by: *i*) placing the audio source in a spatial position that eases his hearing process (e.g., since his right ear is not impaired, sounds coming from his right side could be better perceived); *ii*) incorporating a hearing loss simulator in the 3D audio rendering algorithm, as done in [15]. This customization enables Bob to take off his hearing aid while wearing headphones, thanks to the built-in compensation for his specific auditory impairment. To correctly simulate Bob's hearing loss, the application developers must interact with his otolaryngologist to numerically model and characterize the impairment as accurately as possible.

Given the young age of the user, the assessment of the above-described tool can happen by means of the instructor's observation of Bob's interaction and engagement level. Moreover, Bob's acquired knowledge of various instrumental timbres can be evaluated ex-post, thus indirectly measuring the effectiveness of the tool.

a) 

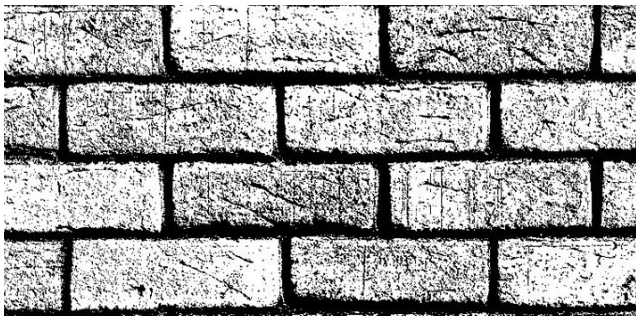
b) 

Fig. 2. Example of *a*) cross-modal conveyance of musical timbre; *b*) black and white image rendering for haptic texture design [14].

C. Third Scenario

Charlie is a 65-years-old orchestra conductor. A few months ago, he suffered a stroke that compromised his motor abilities: at present, he has limited control on the movements of his right limbs, whereas his left limbs did not suffer any damage.

Charlie is currently attending a rehabilitation program, with a tight schedule of physiotherapy sessions that prevents him from resuming in-presence rehearsal and performance activities with his orchestra. However, replicating work-related gestures (e.g., practicing fine motor skills related to orchestra conduction) is a fundamental part of his rehabilitation program, as they are essential for his occupational needs. Therefore, he adopts a customized NMP solution similar to the one depicted in Fig. 1, to enable remote interactions with the orchestra members, who still gather together in their usual rehearsal room. In this setup, gestural data of the hands' joints are not collected via wearable devices but thanks to an Oculus device worn by Charlie. After remotely conveying such gestural data via the NMP system (which should operate in client/server mode to ensure adequate processing capabilities of the gestural data streams, while guaranteeing high-fidelity audio quality thanks to the implementation of PLC mechanisms), they are visualized via a stylized avatar (as the one depicted in Fig. 3) on a screen located in front of every orchestra member. This guarantees scalability to a large number of orchestra components, without need of transmitting a dedicated video stream to each of them. Conversely, Charlie visualizes through his Oculus all the avatars of the orchestra members, with a spatial configuration that replicates their actual placement in the rehearsal room.

In addition to fostering scalability, a further benefit of using an Oculus-based motion tracking system instead of a traditional video streaming is that avatar-based rendering allows for possible manipulations of the acquired gestural data. For example, the movements of Charlie's left hand could be mirrored by the avatar's right hand, thus showing to the remote orchestra both hands replicating the same gestures even if Charlie's right hand is at rest. Moreover, generative models could be exploited to re-create conduction gestures that Charlie is currently unable to correctly make with his right hand. Additionally, specific interactions with the orchestra members could be designed to incentivize the use of the right hand (e.g., a tailored placement of the orchestra components within the virtual space).

Similarly to Scenario A, the assessment of the above-described NMP system can be done by administering questionnaires for Charlie and orchestra members. In addition to all the aspects already enumerated in Section IV.A, the impact of sickness when wearing the Oculus device and the effectiveness of the avatar-based representation should be evaluated.

a) 

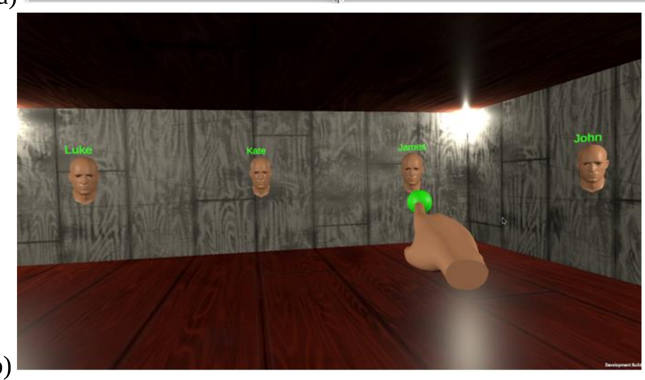
b) 

Fig. 3. Example of avatar-based visualization of remote conductor, currently being developed by the author's research team, including four musicians. The green sphere associated to the conductor's second finger facilitates the visualization of pointing gestures. a) Conductor's avatar pointing at James, seen from the perspective of four virtually connected musicians. b) Conductor's perspective of the four connected musicians.

VI. CONCLUSION

This paper focuses on ensuring equal access to musical education and promoting inclusivity by exploring technological solutions for music students and practitioners facing visual, auditory, or mobility challenges. The proposed interdisciplinary design approach aims to create tailored educational activities for learners with special needs. Illustrative use-cases demonstrate practical applications, in the hope of inspiring further research in inclusive music education.

TABLE I
SUMMARY OF REQUIREMENTS AND CHARACTERISTICS OF THE THREE USE-CASE SCENARIOS

	Scenario A: visually-impaired Conservatory student	**Scenario B: hearing-impaired school pupil**	**Scenario C: motor-impaired orchestra conductor**
General purpose	Real-time remote rehearsals.	Acquisition of the notion of musical timbre.	Real-time remote rehearsals.
Technical requirements	Low-latency, transmission reliability.	Transmission reliability.	Low latency, transmission, reliability, scalability.
Definition of learning requirements: educational goals	Learn gestural coordination with first violinist; adhere to agogic and dynamic indications without reading Braille scores; reinforce accustomation to spatial displacement of quartet members acquired in in-presence rehearsals.	Reinforce acquisition of mental concepts by combining multiple sensorial stimuli (sight, hearing, touch); exploit a serious game as facilitator for conveyance of the notion of instrumental timbre.	Re-learn gestures required for orchestra conduction; activate mirror neurons to reacquire functionalities of the impaired limb by imitation; re-learn how to inspire emotions in orchestra performers via the set of currently reproducible gestures.
Design of engaged learning: interaction requirements	GUI compatibility with screen readers; spatialized audio rendering; remote control of audio settings; shared metronome cues; gestural tracking with haptic feedback.	Superposition of haptic and visual interface; customized binaural audio rendering; adequate storytelling.	Gestural tracking of conductor's movement; gestural rendering via avatars; spatialized audio rendering.
Prototyping: technological components	NMP framework; tactile wearables; binaural audio system.	Tactile display; customized binaural audio system.	NMP framework; motion tracking equipment, binaural audio system.
Evaluation: assessment mechanisms	Questionnaires (e.g., UEQ, SUS, or custom made), Likert-Scale evaluations.	Observation of user interactions.	Questionnaires (e.g., UEQ, SUS, or custom made), Likert-Scale evaluations.

ACKNOWLEDGMENT

This work has been supported by the Italian Ministry for University and Research under the PRIN program (grant n. 2022CZWWKP). The author thanks Prof. Piera Bagnus, Prof. Chiara Nicora and Cristina Greco for their participation to interviews aimed at the definition of the educational goals of the presented use-cases.

REFERENCES

[1] National Center for Education Statistics, U.S. Department of Education, Institute of Education Sciences. (2022) Students with disabilities. condition of education. [Online, Accessed on May 30, 2024]. Available: https://nces.ed.gov/programs/coe/indicator/cgg
[2] C. Rottondi et al., "An overview on networked music performance technologies," *IEEE Access*, vol. 4, pp. 8823–8843, Dec. 2016,
[3] B. Remache-Vinueza, et al., "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, 2021.
[4] M. Giordano et al., Design of vibrotactile feedback and stimulation for music performance, in: *Musical Haptics*, Springer, Cham, pp. 193–214, 2018,.
[5] S. Serafin et al., "Considerations on the use of virtual and augmented reality technologies in music education," in *2017 IEEE virtual reality workshop on K-12 embodied learning through virtual & augmented reality (KELVAR)*. IEEE, 2017, pp. 1–4.
[6] M. Mandanici et al., "The discovery of interactive spaces: Learning by design in high school music technology classes," *Technology, Knowledge and Learning*, vol. 26, no. 4, pp. 1131–1151, 2021.
[7] M. Warburton, et al., "Measuring motion-to-photon latency for sensorimotor experiments with virtual reality systems," *Behavior Research Methods*, vol. 55, no. 7, pp. 3658-3678, 2023.
[8] H. B. Lima et al., "A survey of music visualization techniques," *ACM Computing Surveys (CSUR)*, vol. 54, no. 7, pp. 1–29, 2021.
[9] B. F. Katz and L. Picinali, "Spatial audio applied to research with the blind," *Advances in sound localization*, pp. 225–250, 2011.
[10] L. Turchet et al., "Internet of musical things: Vision and challenges". *IEEE Access*, vol. 6, pp. 61994-62017, 2018
[11] L. Turchet, "Musical Metaverse: vision, opportunities, and challenges". *Personal and Ubiquitous Computing*, vol. 27, no. 5, pp. 1811-1827, 2023
[12] L. Flórez-Aristizábal et al., "Designability: Framework for the design of accessible interactive tools to support teaching to children with disabilities." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, Glasgow, Scotland, pp. 1-16.
[13] R. Battello et al., "Experimenting with adaptive metronomes in networked music performances." *Journal of the Audio Engineering Society*, vol. 69, no. 10, pp. 737-747, 2021.
[14] M, Sacchetto et al., "Collection of Design Directions for the Realization of a Visual Interface with Haptic Feedback to Convey the Notion of Sonic Grain to DHH Students." *2023 4th International Symposium on the Internet of Sounds*. IEEE, 2023.
[15] M. Cuevas-Rodriguez et al.,, "An open-source audio renderer for 3D audio with hearing loss and hearing aid simulations". In AES Convention 142, May 2017.

BIOGRAPHY

Cristina Rottondi (Senior Member, IEEE) is Associate Professor with the Department of Electronics and Telecommunications of Politecnico di Torino (Italy). Her research interests include optical networks planning and networked music performances.

She is co-author of more than 100 scientific publications and co-recipient of the 2020 IEEE Charles Kao award, of the 2022 Journal of the Audio Engineering Society best paper award and of three conference best paper awards (FRUCT-IWIS 2020, DRCN 2017, GreenCom 2014).