Received August 29, 2018, accepted September 24, 2018, date of publication September 28, 2018, date of current version November 9, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2872625

Internet of Musical Things: Vision and Challenges

LUCA TURCHET[1], CARLO FISCHIONE[2], (Senior Member, IEEE), GEORG ESSL[3], (Member, IEEE), DAMIÁN KELLER[4], AND MATHIEU BARTHET[1]

[1]Centre for Digital Music, Queen Mary University of London, London E1 4NS, U.K.
[2]Department of Network and Systems Engineering, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden
[3]College of Letters and Science, University of Wisconsin–Milwaukee, Milwaukee, WI 53211, USA
[4]Amazon Center for Music Research (NAP), Federal University of Acre, Rio Branco 69920-900, Brazil

Corresponding author: Luca Turchet (luca.turchet@qmul.ac.uk)

The work of L. Turchet was supported by a Marie-Curie Individual Fellowship of the European Union's Horizon 2020 Research and Innovation Programme, under Grant 749561, and in part by the Micro Grant of the Marie Curie Alumni Association. The work of C. Fischione was supported by the TNG SRA ICT project TouCHES—Tactile Cyberphysical Networks through the Swedish Research Council. The work of D. Keller was supported by Brazilian Research Council–CNPq through a Productivity Grant. The work of M. Barthet was supported by the EU H2020 Audio Commons Initiative Project under Grant RIA 688382.

ABSTRACT The Internet of Musical Things (IoMusT) is an emerging research field positioned at the intersection of Internet of Things, new interfaces for musical expression, ubiquitous music, human–computer interaction, artificial intelligence, and participatory art. From a computer science perspective, IoMusT refers to the networks of computing devices embedded in physical objects (musical things) dedicated to the production and/or reception of musical content. Musical things, such as smart musical instruments or wearables, are connected by an infrastructure that enables multidirectional communication, both locally and remotely. We present a vision in which the IoMusT enables the connection of digital and physical domains by means of appropriate information and communication technologies, fostering novel musical applications and services. The ecosystems associated with the IoMusT include interoperable devices and services that connect musicians and audiences to support musician–musician, audience–musicians, and audience–audience interactions. In this paper, we first propose a vision for the IoMusT and its motivations. We then discuss five scenarios illustrating how the IoMusT could support: 1) augmented and immersive concert experiences; 2) audience participation; 3) remote rehearsals; 4) music e-learning; and 5) smart studio production. We identify key capabilities missing from today's systems and discuss the research needed to develop these capabilities across a set of interdisciplinary challenges. These encompass network communication (e.g., ultra-low latency and security), music information research (e.g., artificial intelligence for real-time audio content description and multimodal sensing), music interaction (e.g., distributed performance and music e-learning), as well as legal and responsible innovation aspects to ensure that future IoMusT services are socially desirable and undertaken in the public interest.

INDEX TERMS Internet of Things, sound and music computing, ubiquitous music, mobile music, networked music performance, participatory art.

I. INTRODUCTION

The umbrella term "Internet of Things" (IoT) broadly refers to the extension of the Internet into the physical realm, by means of everyday physical objects that are spatially distributed and augmented using information and communication technologies [3], [23], [98]. In the Internet of Things, "Things" refer to embedded systems that are connected to the Internet, which are able to interact with each other and cooperate to reach common goals. Things are characterized by embedded electronics, wireless communication, sensing, and/or actuation capabilities.

The manifestation of the IoT has led in recent years to a substantial increase in smart devices and appliances in the home, office, cities, and other environments that connect wirelessly through local networks and the Internet. More and more daily objects are becoming embedded with sensors, actuators, processing elements, and are gaining the ability to communicate wirelessly. To date, however, the application of IoT technologies in musical contexts has received remarkably little attention compared to other domains such as consumer electronics, healthcare, cities, and geospatial analysis.

Recently, the term "Internet of Musical Things" (IoMusT) has been used by various authors with different perspectives. Hazzard *et al.* [58] used the term within the context of musical instruments augmented with QR codes pointing users towards online data about the instrument and its social history. Keller and Lazzarini used the term within the context of ubiquitous music (ubimus) research [72], which encompasses creative usage of the Internet, of mobile devices, and of embedded technologies for ubiquitous musical activities [73]. Turchet *et al.* [140] proposed extending the concept of IoT to the musical domain, considering IoMusT as a subfield of IoT where the underlying technological infrastructure enables ecosystems of interoperable devices connecting musicians and audiences, to support novel musician-musician, audience-musician and audience-audience interactions. Roy *et al.* [115] used the term in the context of crowdsensing for musical applications, proposing an architecture to address the efficient use of sounds produced by a large-scale community.

The main purpose of this paper is to provide a unified vision of the IoMusT paradigm. We will focus on technologies enabling the IoMusT as well as on current IoMusT research activities, drawing attention to the most significant contributions and solutions proposed over the recent years. Currently, active research on IoMusT-related themes is rather fragmented, typically focusing on single technologies or single application domains. Such a fragmentation is potentially detrimental for the development and successful adoption of IoMusT technologies. Consequently, this article not only seeks to bridge existing research areas and communities and foster cross-collaborations, but also aims to ensure that IoMusT-related challenges are tackled within a shared system-level perspective. The industrial perspective also has a very important role in the success of the IoMusT paradigm. Therefore, we will discuss examples of technologies developed by various music hardware and software manufacturers, which provide an indication of mid- and long-term industrial priorities.

The IoMusT may foster new opportunities for the musical industry, paving the way to new services and applications capable of exploiting the interconnection of the digital and physical realms. Nevertheless, for the IoMusT technologies to emerge and be adopted by end users a number of technical, artistic, and pedagogical challenges need to be addressed. These include low-latency communication infrastructures and protocols, embedded IoT hardware specialized for audio, dedicated application programming interfaces (APIs), software relying on ontological and semantic audio processes, and the design of novel devices dedicated to music production or consumption. This paper aims to identify and discuss such challenges.

The remainder of this article is organized as follows. Section II introduces the conceptual basis of the IoMusT vision. Section III surveys works and technologies related to the envisioned IoMusT. Section IV sketches a set of hypothetical IoMusT scenarios, and uses them to identify key capabilities missing from current systems. Section V discusses the main research challenges ahead of us on the IoMusT landscape that need to be tackled to develop these capabilities. Finally, we provide summarizing conclusions in Section VI.

II. THE INTERNET OF MUSICAL THINGS: CONCEPT AND VISION

The Internet of Musical Things is an emerging field positioned at the intersection of Internet of Things [3], [23], [98], new interfaces for musical expression [65], ubimus [73], networked [113] and mobile music [44], [134], human-computer interaction [110], [114], participatory art [161], and artificial intelligence applied to musical contexts [26].

From a computer science perspective, we define a Musical Thing (MusT) as *"a computing device capable of sensing, acquiring, processing, or actuating, and exchanging data serving a musical purpose"* and the IoMusT as *"the ensemble of interfaces, protocols and representations of music-related information that enable services and applications serving a musical purpose based on interactions between humans and Musical Things or between Musical Things themselves, in physical and/or digital realms. Music-related information refers to data sensed and processed by a Musical Thing, and/or exchanged with a human or with another Musical Thing"*.

The IoMusT has strong connections with and could be seen as a subfield of the Internet of Media Things, which is currently under exploration by MPEG.[1] Similarly to what the Web of Things[2] represents for the Internet of Things, we use *"Web of Musical Things"* to refer to approaches taken to provide an Application Layer which supports the creation of Internet of Musical Things applications.

Just like the general IoT field, the IoMusT may encompass manifold ecosystems [21], [91]. An *IoMusT ecosystem* is composed of users involved in musical activities (e.g., performers, audiences), information and service providers, and forms around commonly used IoMusT hardware and software platforms as well as standards. From the technological perspective, the core components of an IoMusT ecosystem are of three types:

(i) Musical Things: The IoMusT vision predicts that in the future, a new class of musical devices will be connected to the Internet, which could have a transformative effect on how humans involved in musical activities (e.g., performers, audiences) conduct these activities and interact with musical objects (e.g., musical instruments). We position Musical Things as a subclass of Things, therefore they inherit characteristics of Things, such as sensors, actuators, connectivity options, and software to collect, analyze, receive, and transmit data. Key factors are interoperability and synchronization. Musical Things are entities that can be used in a musical context to produce musical content or to observe phenomena associated to musical experiences, and can be connected to a local and/or remote network and act as sender and/or receiver. A Musical Thing can be, for example, a smart instrument [143], a musical haptic wearable [138], or any

[1] ISO/IEC 23093 (IoMT): https://mpeg.chiariglione.org/standards/mpeg-iomt

[2] https://www.w3.org/WoT/

TABLE 1. Examples of technological components of an IoMusT ecosystem.

Component	Examples
Musical Things	Smart instruments; musical haptic wearables; networked speaker systems; intelligent mixing consoles; networked virtual reality headsets;
Connectivity	Wireless and wired networks supporting, in both local and remote communications, ultra-low latency, high reliability, high throughput, high quality of audio and multimodal content; standards and protocols for the IoMusT (e.g., supporting synchronization mechanisms); antennas specific to IoMusT networks; Web of Musical Things; APIs (based on IoMusT API specifications);
Services	Service for connecting smartphones to smart instruments; service for connecting smart instruments to social networks; service for creative content analysis; service for cross-modal mappings between sensed data and control parameters of Musical Things; service for content synchronization between Musical Things;
Applications	For audience members: enhanced concert experiences based on multisensory content provided by Musical Things, remote audience participation in the music creation process during concerts; for musicians: remote rehearsals, interaction with the cloud directly from smart instruments; for audio engineers: intelligent live and studio production supported by smart instruments; for students: enhanced music e-learning with remote teachers, or augmented experiences through mobile apps and mixed reality displays leveraging data from smart instruments; for music teachers: web-based apps for student assessment, progress monitoring, and individual or group feedback (thanks to the smart instrument capabilities).

other networked device utilized to control, generate, or track responses to musical content [44].

(ii) Connectivity: The IoMusT connectivity infrastructure supports multi-directional wireless communication between Musical Things, both locally and remotely. The interconnection of Musical Things over local networks and/or the Internet is achieved by means of hardware and software technologies, as well as standards and protocols governing the communication. Music performance puts particular constraints on communications. Specifically, in typical real-time use cases the connectivity infrastructure should ensure communications with low latency, high reliability, high quality, and tight synchronization between connected devices. These requirements are similar to real-time tele-surgery, virtual and augmented reality, and real-time automation, but are not required in the vast majority of IoT applications.

(iii) Applications and Services: Various types of applications and services can be built on top of the connectivity, which target users such as musicians (e.g., composers, performers, conductors), audio engineers (e.g., live or studio sound engineers) and audience members. Such applications and services may have an interactive or a non-interactive nature. To establish interactive musical applications, real-time computations have a particular importance. Analogously to the IoT field, the IoMusT can leverage Web APIs and Web of Things architectures [55] designed to serve musical purposes. Services can be exposed by Musical Things via Web APIs. Applications are part of a higher layer in the Web of Musical Things architecture letting users interact with content or Musical Things directly. To illustrate the novelty of the IoMusT, we briefly discuss here examples of services and applications that musicians could have access to in the future. For example, the IoMusT could support services letting users connect to smart musical instruments via web technologies. This could let fans follow what and how a musician has been playing and where. This could also benefit music learners through data analytics characterizing playing style (e.g., playing frequency, mistakes, patterns, or tunes recently played). Smart musical instruments could be connected to social media platforms such as Facebook or Twitter letting musicians share information about what they play and when, leading to novel musician-fan relationships.

Table 1 provides a non-exhaustive list of examples for each of the technological components of an IoMusT ecosystem.

The IoMusT infrastructure enables ecosystems of interoperable devices connecting musicians with each other, as well as with audiences, which multiplies the interaction possibilities both in co-located and remote settings (see Figure 1). Such ecosystems provides fertile grounds for the design of creative artifacts providing novel types of affordances, including ways to monitor creative control or response to musical content associated to musicians and/or audience members. By applying the IoT field to the musical domain we envision a departure from the traditional Western written music communication chain (i.e., composers writing musical content interpreted by instrumentalists, and received by a passive audience) towards a musical mesh where possibilities of interactions are countless. We envision both small scale (e.g., co-located music performance in a concert hall) and large scale scenarios (e.g., massive open online music performance gathering thousands of participants in a virtual environment).

Figure 2 shows a conceptual diagram of the different components that are interconnected in our vision of an IoMusT ecosystem. As can be seen in the diagram, interactions between human actors are mediated by Musical Things. The diagram shows interactions between performers and audience members but it can be further extended to include other actors, such as live sound engineers, studio producers, composers, conductors, or educationalists. The interactions can be both co-located (see blue arrows), when the participants are in the same physical space (e.g., concert hall, public space), or remote, when they take place in different physical spaces that are connected by a network (see black arrows).

Regarding co-located interactions, these may be based on point-to-point communications between two Musical Things, but also on broadcast communications between multiple Musical Things (see the blue dashed arrows). Examples of these one-to-one and many-to-many interactions are the delivery of content (e.g., such as text or visuals) from the

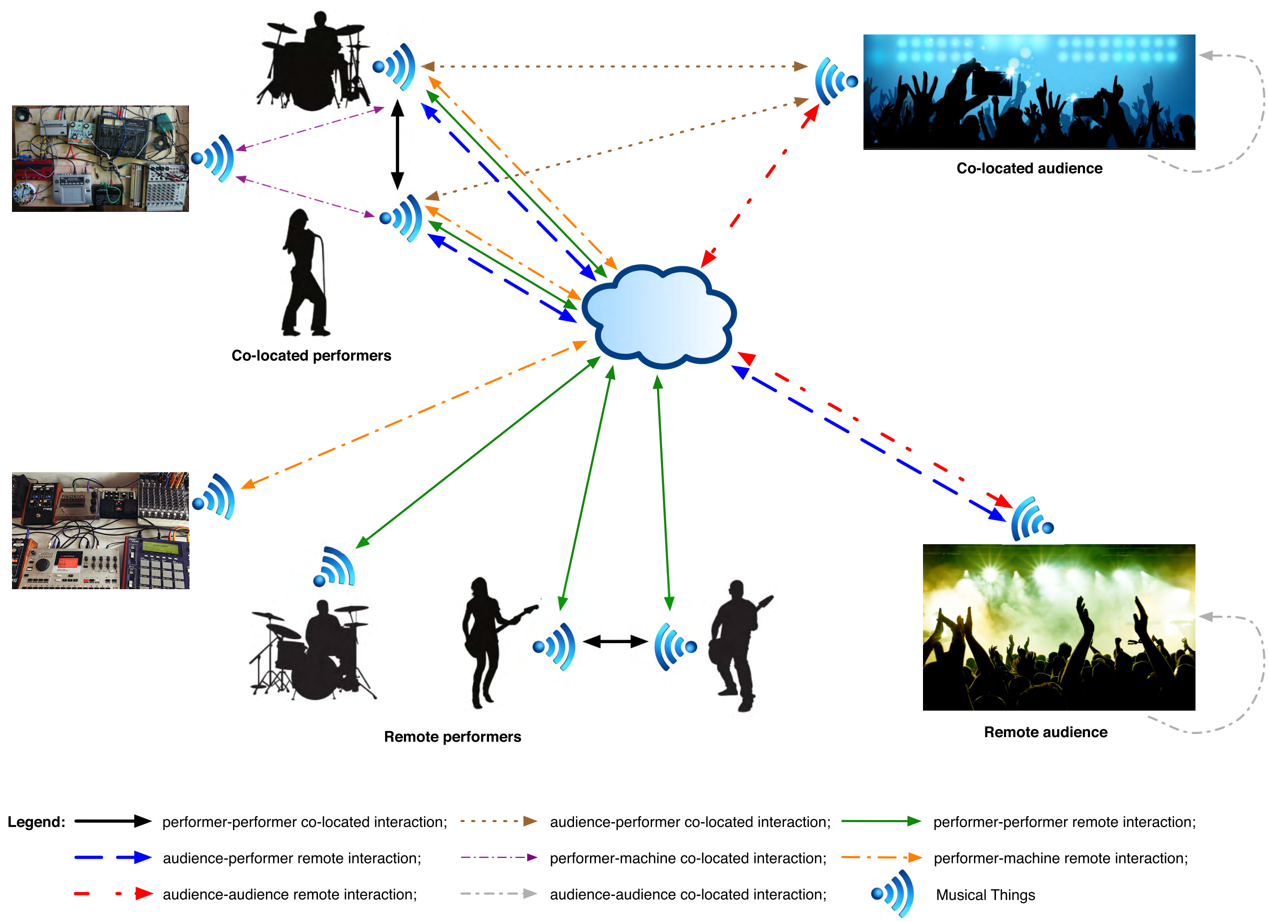

FIGURE 1. Interaction possibilities between musicians, audience members, and machines.

musicians' smart musical instruments [143] to the audience members' mobile phones, and vice versa. Co-located interactions might also occur at the collective levels for both musicians and audiences (see the blue solid line arrow). An example of the latter case may involve one or more smart musical instruments affecting the smart lighting system of a music venue. Moreover, Musical Things may offer novel ways to interact with musical content for both musicians and audience members. For instance, a smart musical instrument could recommend a set of songs to a musician for pedagogical purposes.

Regarding remote interactions, these may occur between audience members/musicians present at the concert venue and remote audience members/musicians (see the solid black arrows). They may also occur between remote audience members/musicians (see the black dashed arrows). Such remote interactions could be delivered through virtual reality systems with audiovisual communication channels.

Methods of communication between Musical Things would be defined in APIs (indicated in Figure 2 with the small red rectangles) based on a unified *IoMusT API specification*. The interactions mentioned above, based on the exchange of multimodal creative content, are made possible thanks to a set of *services* (indicated with the green areas). For instance, these can be services for creative content analysis (for example, based on multi-sensor data fusion [57] and music information retrieval [26]), services for creative content mapping (between analysis and devices), or services for creative content synchronization (between devices). Multimodal mapping strategies are required to transform, especially in real-time, the sensed data into control data for generation purposes (e.g., haptic, auditory, and visual content).

III. RELATED FIELDS

In this section we review key works related to the IoMusT vision. The review is not meant to be exhaustive, rather we aim to describe the results of various application domains that lead to the emergence of the IoMusT field.

A. NETWORKED MUSIC PERFORMANCE SYSTEMS

The notion of musical performances over networks is almost as old as the networks themselves. Connected music performances go back to at least 1978 when the League of Automatic Music Composers connected three computers to exchange performance-centric information [149]. The emergence of the Internet brought a drastic expansion of the

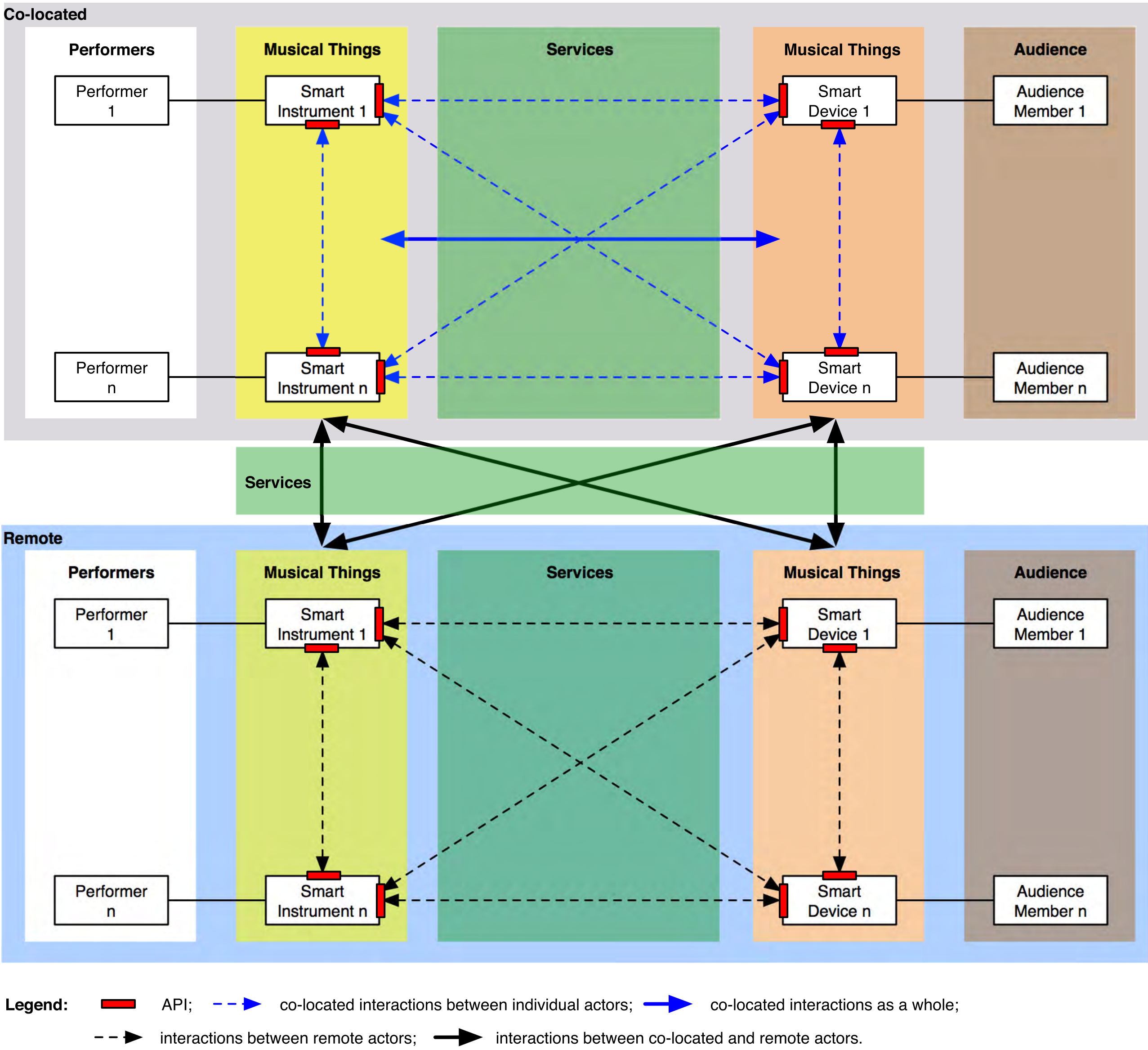

FIGURE 2. Block diagram of an IoMusT ecosystem involving performers and audience members.

potential for networked music performances (NMPs). This led to a vibrant engagement with the creation of NMP systems [9], [53], [113], [150]. In [113], Rottondi *et al.* provide a comprehensive overview of hardware and software technologies enabling NMPs, including low-latency codecs, frameworks, protocols, as well as a discussion of perceptually relevant aspects. A notable example of NMP system is the ReacTable [68], a tangible interface consisting of a table capable of tracking objects that are moved on its surface to control the sonic output. The ReacTable allows multiple musicians to simultaneously interact with objects either placed on the same table or on several networked tables in geographically remote locations.

Networked collaborative music creations can occur locally and globally. They can occur either over wired networks such as Local Area Networks (LANs) or Wide Area Networks (WANs), or over wireless networks such as Wireless Local Area Networks (WLANs) [53] or global cellular data networks [44]. Furthermore, local things can be interconnected via Body Area Networks (BANs) [104] or Personal Area Networks (PANs) [67] enabled through a portable hub such as a mobile device. Currently, networking has achieved in urban centers a kind of ubiquity such that availability is usually no longer the concern. Rather, the problems to address are related to enabling music performance and overcoming constraints. Important interrelations between desirable features for music performance and network constraints can be found in research activities on the following topics:

(i) Low latency. NMPs require low latency to provide conditions that are similar to those in natural music making settings where sounds produced by various musicians can be heard almost instantaneously due to the fast propagation of acoustical waves in airs. Substantial research has been conducted to find solutions to minimize latency and its variability while maximizing audio quality [113]. One such example is the JackTrip streaming system [27]. Regarding NMPs over WLAN, a set of recommended Wi-Fi configurations to reduce latency and increase throughput for live performance scenarios has been proposed in [101].

(ii) Synchronization. Empirical works on delay induced in musical performance indicate that long delays interfere with synchronization among musicians, but a range of 10-21ms may offer natural playing conditions. Latencies of 25-60ms allow synchronization, albeit with a trend of decelerating tempo. Above 65ms, the rhythmic aspects of performance have been found to deteriorate [34]. The presence of deceleration drift suggests that persistent metronome pulses can help synchronization in networked performances [29]. One technique is to create a feedback signal such as a metronome pulse that is adjusted dynamically to the current network delay. The usefulness of this approach has been demonstrated in a network performance between Palo Alto (USA) and Beijing (China) with an average unidirectional latency of 110ms [28]. Another approach seeks to generate a globally shared synchronized time by incorporating GPS hardware into networked music devices [103]. In the context of web audio technologies, a synchronization solution based on HTML5 has been proposed using a reference time shared across devices [78]. This solution was effective in achieving the distributed rendering of audio events with an accuracy of 1 to 10ms, and 5ms in standard deviation (no audio streaming was involved, only the exchange of control information as pre-composed content was stored on the client side). Regarding industrial contexts, the company Ableton developed Link,[3] a synchronization technology that keeps software applications in time over both LAN and WLAN. Link provides synchronizations based on musical beat across multiple applications running on a single or multiple devices. Such applications discover each other automatically within the LAN/WLAN and any musician using them can change the tempo as well as leave or join without disrupting the session.

(iii) Interoperability through standardization. Open Sound Control (OSC) is a flexible protocol for networked interoperability of musical control information [120], [160]. One of the features of OSC is its open address space structure that allows for easy incorporation of new musical services over a network. For this reason OSC has been used as the underlying communication protocol in applications involving smart devices, for example musical uses of smart watches [97]. On the other hand, to date, Link appears to be a candidate for becoming a standard as far as music synchronization mechanisms are concerned. One of the reasons lies in the fact that its source code has been freely released according to the GNU General Public License, which is leading to adoption of the technology by more and more music software manufacturers and developers in general.

(iv) Seamless integration and ease of participation. OSC itself, however, does not provide suitable standardizations of functionality, these instead have to be defined by developers [51]. This suggests that some level of standardization of OSC name spaces for NMP could be helpful to allow musical devices to be more seamlessly integrated. Furthermore, seamless integration advances the ease of participation for musicians, audiences or other individuals involved in the music production chain, in NMPs. Alternatively, information sharing mechanisms that offer ways to discover local name spaces and link them to semantic meanings could potentially address this issue. One proposal along these lines is to use service discovery and "zero configuration" mechanisms of networks for information exchange [43], [88]. This approach allows participants to join and leave ongoing NMPs by discovery. Ease of participation can also involve the local distribution of performance software. Discovery and peer-to-peer Wi-Fi can be used in this context to support such solutions from a network protocol level. MaNet is one such example of a technology that offers mechanisms to minimize the burden of distribution and configuration in audience participation pieces [83].

(v) Scalability. One can envision NMPs that have very large and extremely distributed participation. Such performances put particularly strong burden on the desirable properties discussed so far. Cloud computing, which refers to the utilization of large computational capacity available over networks, can be used to support scalability in NMP [2]. For example, cloud computing can be used to off-load and scale sound synthesis computations [59]. It can also be used to scale data distribution through cloud-based scalable message distribution [82]. Evidence for performance characteristics and scalability was collected as part of the "Crowd in C [loud]" mobile audience participation piece. Performance was evaluated with a scaling audience participation that peaked at 48 concurrent mobile devices [39], still a relatively small number.

B. IOT TECHNOLOGIES

The background communication technologies of the IoMusT are Wireless Sensor Networks (WSNs) [38], Internet of Things (IoT) [3], [23], [98], and Tactile Internet [1], [48].

WSN refers to a network of nodes that can communicate via wireless networks for purposes such as monitoring, communication, and automation. The nodes of a WSN are ultimately electronic devices that can be embedded potentially in any physical object. The term Internet of Things later emerged, proposing that WSNs could be reached via the Internet Protocol. Given their high societal impact, these technologies have been the object of intense research in many disciplinary domains, ranging from Electrical Engineering to Computer Sciences, and both in academia and industry (see e.g., [156]). Within communication networking engineering, this has resulted in the emergence of new communication protocols, especially for low data rate and low

[3] https://www.ableton.com/en/link/

power consumption, such as IEEE 802.15.4 [37], Zigbee,[4] and ROLL.[5] While initially WSNs and IoT were mostly targeting local and personal area networks, more recently researchers have been considering the integration of WSNs and IoT together with cellular wireless communication systems in the so called wireless networks of 5-th generation (5G networks). Within 5G, large companies such as Ericsson are proposing the standardization of networking protocols that will be capable of connecting the objects of a WSN or of IoT via cellular wireless base stations, the same stations that we use today for our cellular phones. The state-of-the art of this activity is the narrow-band IoT [79].

Despite the theoretical and technological advancements described above, unfortunately, most of the existing wireless communication protocols of WSNs and IoT cannot be used for applications relying on the wireless networking of musical instruments that require ultra-low latency and high reliability. Although most cutting-edge networks of 5G will deliver very high data rate thanks to the usage of millimeter Waves (mmWaves) communication [123], [124], they will also provide communication latencies on the order of tens of milliseconds [79]. This is generally insufficient for the wireless interconnection of musical instruments. To transmit music streams from one instrument to a listener located on another side of the communication network, communication latencies on the order of milliseconds are needed [112], [113], while the probability of unsuccessful message receptions has to be on the order of 10^{-10} to avoid perceivable deteriorations of the signal [48]. These stringent requirements are not only required by the IoMusT, but also by a plethora of future technologies, such as virtual reality, telepresence, telesurgery, autonomous car driving, and smart power transmission grids to mention a few. Such a large class of technologies demanding low latency communications with high reliability is motivating the development of the emerging paradigm of the *"Tactile Internet"* [1], [48].

What is missing to the state-of-the-art technological advancements of WSNs and IoT described above is the real-time dimension [66]. Technologies such as the IoMusT can be groundbreaking only if there are reliable communication protocols with low latencies. To address such a challenge, Tactile Internet research proposes to substantially augment an Internet network such that the communication delay between a transmitter and a receiver would be so low that even information associated to human touch, vision, and audition could be transmitted back and forth in real-time with regard to the human senses. This will enable seamless remote interaction experiences. The Tactile Internet is not only expected to ensure low latencies with very low probability of missing messages, but also to enable both low and high data rates. Within the wireless part of the communication channels, this will be possible via wireless transmissions over the mmWaves frequencies. These are wireless frequencies within the range of 10 to 300 GHz. MmWaves communications will offer data rates of Giga bits per second over relatively short distances. Within the realm of musical instruments, such a wireless technology appears particularly interesting due to the very small size of the antennas and therefore of the transceivers. Thus the electronic platform supporting such communications will be easily embedded in musical instruments (as proposed in the Smart Instruments concept [143]). Moreover, the high data rates will enable the transmission of multimodal content in high resolution.

While the wireless transmission medium is the most critical from the point of view of ensuring low latencies over relatively limited geographic areas on the order of some kilometers, the wired media is the one imposing ultimate limitations on the geographic distance over which the communications will be possible [48]. After the messages are transmitted over the wireless medium, they are forwarded over some wired medium such as copper cables, or more likely, optical fibers. If we make the ideal assumption that only optical fibers are used, the speed of light is the fastest speed at which messages can be theoretically transmitted. Therefore, point-to-point communications with latencies of around 1ms would be realizable among performers that are spread within a distance of at most 300Km.

C. INTERACTIVE PERFORMANCE

Interactive performance commonly refers to practices that offer sensor-equipped live performers (e.g., dancers or musicians) real-time control over media elements and events using their physical movements and gestures [102]. With the advent of digital musical instruments for which the mechanical gestural actions from performers are severed from natural acoustical effects, there is an infinite number of possibilities to map control attributes to perceptual ones. As such, in digital musical instruments design, mapping techniques represent a central challenge which is addressed within the New Interfaces for Musical Expression endeavor. Malloch *et al.* [87], [88] proposed libmapper, a 3-layer mapping framework and tools where a "semantic layer" links gestures to sound semantics. This gesture-instrument mapping paradigm was then extended to more general issues of sharing and manipulating multiple data streams among different media systems in heterogeneous interactive performance environments (SenseWorld DataNetwork) [5]. The libmapper tool and the SenseWorld DataNetwork rely on OSC. The libmapper technology provides decentralized resource allocation and discovery, and flexible connectivity letting devices describe themselves and their capabilities. However, it targets the use of a LAN subnet where support for multicast can be guaranteed [89].

Another framework supporting sensor-based interactive performance is Camurri *et al.*'s EyesWeb XMI which seeks to extend music language toward gesture and visual languages, with a focus on analyzing expressive content in gesture and movement, and generating expressive outputs [145]. EyesWeb is more oriented towards theatre and dance performing arts with a particular focus on movement qualities

[4] www.zigbee.org
[5] www.ietf.org/dyn/wg/charter/roll-charter.html

inferred from computer vision techniques. In a different vein, the Social Signal Processing framework [146] supports the development of online recognition systems from multiple sensors. Its architecture is established to handle diverse signals (e.g., audio, heart beat signal, or a video image). These frameworks were used in ''soft'' real-time interactive performance applications where changes are relatively slow (e.g., theatre, art installations). Extending them to musical interactions within an IoMusT ecosystem would require to tackle interoperability and delay challenges, as discussed in Section V-A.

D. UBIQUITOUS MUSIC

Ubiquitous music (ubimus) refers to music or musical activities that are supported by ubiquitous computing concepts and technology [118], [151]. It has been defined as *"ubiquitous systems of human agents and material resources that afford musical activities through creativity support tools"* [73]. Ubiquitous music research has contributed to the advancement of IoMusT-based creative proposals.

As an attempt to establish a consensual definition of ubimus practice, Keller and Lazzarini [72] discuss four components of ubimus activities. Component 1, linked to human-related aspects, finds theoretical foundations in emerging views on evolutionary theory. Given their reliance on social interaction, ubimus systems can be construed as *behavioral ecologies*. These ecologies foster and are shaped by human behavior. Ubimus projects that make intensive use of environmental features to modulate the behaviors of the agents or to generate and organize material resources emphasize the role of component 2 of the ubimus definition - the material resources. Given their heavy usage of environmental features and the exploration of the emergent qualities of behavioral patterns across modalities, ubimus systems can be described as *multimodal ecologies*.

The third component of the definition relies on interactions between components 1 and 2, encompassing *relational properties*. Three categories of relational properties have been proposed: material relational properties, which arise from activities with physical objects; social relational properties, which emerge from exchanges among the stakeholders; and formal relational properties, which are featured in cognitive simulations and conceptual operations that employ off-line cognitive resources - i.e., resources that are decoupled from the activity [74], [158].

The fourth component of the definition of ubiquitous music is *ubimus ecosystems*. Ubimus ecosystems function as technological hubs that support the integration of audio and interaction tools. These ecosystems can be reconfigured according to the needs of the users through rapid prototyping techniques. As a case study, Lazzarini *et al.* [80] report the development of PNaCl Csound, which provides an environment that can be employed for the development of a variety of ubimus applications on standard Internet browsers. It is based on a domain-specific programming language - Csound - that features a wide-ranging variety of sound generators. Hence, it supports the prototyping of reasonably complex audio processing algorithms with a relatively low-latency performance. Furthermore, the Csound PNaCl environment incorporates the know-how developed over thirty years of Csound usage, providing a path for the development of ubiquitous music ecosystems based on standard web browser technology. On the down side, a limitation shared by all browser-based software development systems is their dependence on support and updates of browser technology.

The integration of multiple technological objects into ubimus ecosystems opens new opportunities for artistic applications of the Internet of Musical Things. Small computing units can be remotely controlled to gather data and to interact with people and with material resources. Combinations of customized hardware and ready-made components can foster the integration of computing devices and peripherals. As an example of the artistic applications of the IoMusT, the Memory Tree project [108] uses simple recording and playback devices deployed at a remote physical location but accessible through the Internet. This multimodal installation lets users record short audio excerpts via a social-network tool. The sonic snippets are made available for others to listen to at the installation site, featuring a tree with playback devices. The artistic proposal probes the usage of mobile phones and an Internet-service infrastructure (for the production and deployment of content), featuring custom-made, do-it-yourself hardware (for the on-site playback system). In such a IoMusT scenario, location and environment are tightly integrated with the ubimus ecosystem. As observed in projects such as the Memory Tree, ubimus ecosystems may support both local and remote forms of social interaction. IoMusT functional extensions can increase the geographical and the social significance of ubimus activities, fostering community engagement beyond co-located support. They also provide means of spreading the computational load of ubimus interventions over a heterogeneous collection of units.

An important feature of the ubimus frameworks is their focus on high-level conceptual and methodological approaches. Hence, ubimus proposals usually target concepts that do not depend on specific implementations. Another characteristic is a strong reliance on empirical methods, targeting a slowly growing body of evidence. These views foster the development of small-footprint technology for music making in everyday settings. Consequently, the deployment of the IoMusT may enhance the opportunities for massive ubimus deployments without demanding high investment in resources.

E. WEB AUDIO, CLOUD COMPUTING, AND EDGE COMPUTING

The Web Audio, cloud and edge computing technologies we review in this section are promising technologies to enable scalable audio-based interactive applications in the context of a Web of Musical Things.

Web Audio is a high-level JavaScript API for processing and synthesizing audio in web applications.[6] It represents a step forward compared to previous audio technologies for the web such as plugins (Flash, QuickTime), or the HTML5 audio element that allows for basic streaming audio playback. Web Audio supports some of the features found in modern digital audio workstations such as the routing of audio signals, high dynamic range, sample-accurate sound playback with low latency, mix processing (e.g. dynamics- and delay-based effects, equalization, spatialization), as well as synthesis. Recent works have used Web Audio for technology-mediated participatory live performances involving mobile phones. Examples include [105], which proposes a framework allowing composers to control web-based audio processes on smartphones of audience members who can participate creatively by modulating the sound through sensor-based interactions. The system reported in [129] aims to engage the audience in music making by sonifying messages from a multi-user chat system using Web Audio synthesis. Other works bring digital audio workstation features to the web from virtual instrument amplifiers [25] and sound effects [6] to a collaborative music production platform [85].

For musical applications requiring the exchange of real-time data, the WebSockets technology[7] provides full-duplex communication channels over a single TCP connection enabling a client to send messages to a server and receive event-driven responses without having to poll the server for a reply. Although WebSockets require a client/server architecture, a more recent initiative, WebRTC,[8] allows browser-based audio and video communication through direct peer-to-peer communication. The use of Web Audio over WebRTC is hence an interesting avenue to facilitate collaborative musical interactions directly from the browser.

Cloud computing structures rely on an abstraction of distributed servers in order to simulate a centralized network enabling load balancing and resource replication. As proposed in [33], services deployed on a cloud computing structure may be used to facilitate the intercommunication between distributed musical applications in a live performance context by minimizing the amount of network configuration necessary, and improving scalability and message reliability. The push notification cloud services provide ways to deliver messages through web sockets and HTTP streaming, saving the overhead of restarting TCP connections each time new data are sent between a client and a server (like for the HTTP REpresentational State Transfer (REST) architectural style). It is possible to use such push services for long distance networked music by using lightweight messages that contain symbolic data or control signals interpreted by client applications to produce sounds using synthesis [33], [39]. Cloud computing can benefit the IoMusT by providing services that require vast amount of computational power and scalability, going beyond the capabilities of Musical Things' embedded systems (e.g., a multi-user computational musical creativity system enabled by complex artificial intelligence techniques). Cloud computing technologies can also allow music software applications to leverage third party intelligent music services (e.g., an Internet-connected digital audio workstation capable of interacting with smart musical instruments and the cloud for music e-learning and intelligent production applications).

Edge computing is a paradigm that combines desirable properties of cloud computing with real-time requirements [22], [62], [90], [119], [122], [162]. Edge computing refers to the set of technologies that allow computation to be performed at the edge of the network. The term "edge" relates to computing and network resources along the path between data sources and cloud data centers [122]. Pushing the computation towards the edge of the network, where data is produced, offers several benefits compared to the cloud computing paradigm. Firstly, the latency between a device and a service requested by it is kept minimal, because by exploiting the computational resources of the edge it is possible to offload part of the workload from the cloud thus saving the data transmission time (whereas in the cloud computing paradigm, the majority of the processing, such as collecting information and making decision, happens in the centralized cloud). Secondly, the bandwidth between the edge and the cloud can be saved since the data is preprocessed at the edge and therefore the upload data size will be significantly reduced. Thirdly, the edge energy efficiency may be improved thus increasing its battery life. Indeed, the wireless communication module of an edge is usually very energy hungry, therefore offloading some computations to the edge may limit the need of transmitting data and, as a consequence, decrease the usage of such module. These benefits make edge computing a paradigm well-suited to support networked real-time applications [56] and IoT ecosystems [22], [132], such as those envisioned in the IoMusT. A recent endeavor in the application of edge computing techniques to the musical domain is represented by the work of Roy *et al.* [115]. The authors describe a music crowdsensing architecture that optimizes transmission and service time, power consumption, and service energy.

F. TECHNOLOGY-MEDIATED AUDIENCE PARTICIPATION

Technology-mediated audience participation (TMAP) [69], [161] is a fertile field in interactive arts capitalizing on information and communication technologies with the promise of democratizing access to music making and increasing the active engagement of audiences in live music performances. Reviews of TMAP systems can be found in [61], [82], [134], and [161]. These systems disrupt the traditional unidirectional chain of musical communication for written Western music, where the musical messages are exchanged sequentially from composers to performers to listeners. In performances with TMAP, rather than staying creatively "passive", audiences are actively engaged in the

[6] https://www.w3.org/TR/webaudio/

[7] https://developer.mozilla.org/en-US/docs/Web/API/WebSockets_API

[8] https://webrtc.org/

music making process. Certain designs establish mechanisms for audience-performer interactions, such as Mood Conductor [46], which invites audiences to conduct performers through dynamic votes of musical moods, Open Symphony [161], [163], which prescribes audience members a collective role in determining the musical structure of a live piece, and A.bel [36], which lets multimedia artists distribute interactive content onto audiences' mobile devices. Other designs turn the crowd into composers by letting them generate musical motifs on their mobile phones [44], [82], [152], or provide ad-hoc tangible interfaces for co-creation [16], [52]. Most recent TMAP systems make use of Web Audio (see Section III-E).

The majority of current TMAP systems require the audience to use a single type of device and application. Nevertheless, using different types of devices simultaneously could further enrich interaction possibilities. To date, audience creative participation has mainly been based on manual controls or gestures using smartphones (e.g., screen touch, tilt) [44]. Expressive modalities could be increased by tracking physiological parameters (e.g., electrodermal activity, heart rate) [7], [133] at the individual and collective levels using devices specifically designed for this purpose, or by tracking more complex audience behaviors and body gestures. Furthermore, means of interaction in current TMAP systems typically rely on the auditory or visual modalities, while the sense of touch has scarcely been explored to create more engaging musical experiences [45].

G. WEARABLES

The last decade has witnessed a substantial increase in wearable devices in the market. These are smart devices that can be worn on the body as accessories, and are capable of tracking body activity (e.g., gestures, body temperature, galvanic skin response, heart rate) as well as wirelessly exchanging data through Internet or point-to-point connections with other smart devices. In some cases, a small display, speaker or tactile actuator may be included. A distinctive feature of such devices is their unobtrusiveness: they are designed to be worn during everyday activity and to passively collect data without regular intervention by the user. This characteristic makes these devices suitable to track and collect body movements and physiological responses of audience members, for example during live concerts. However, to date, this challenge has been scarcely addressed. A noticeable exception is the work reported in [97] where the accelerometers embedded in smart watches are utilized to track gestures of the user and are mapped to the parameters of a sound engine by leveraging a smartphone as a bridge. Another example is the work reported in [131] where electrodermal activity gauging listeners' arousal is used to generate a visual accompaniment to music.

Recently, a novel class of wearable devices, the *musical haptic wearables*, has been proposed, targeting both musicians and audience members [138]. These devices may encompass haptic stimulation, tracking of gestures and physiological parameters, and wireless connectivity features. Musical haptic wearables were conceived to enhance creative communication between musicians as well as between musicians and audience members by leveraging the sense of touch, in both co-located and remote settings. They were also devised to enrich musical experiences of audiences of music performances by integrating haptic stimulations, as well as provide new capabilities for creative participation thanks to embedded sensor interfaces. An example of musical haptic wearable is Vibropixel, which has been used to assist a conductor with a tactile representation of metronome clicks [63].

In a different vein, recent years have also seen the emergence of electronic textiles (e-textiles), which consist of garments enhanced with fabric-based sensors [153]. This technology may provide some advantages compared to other wearable devices, such as more comfort, an even smaller obtrusive character, and a natural interface for human interaction, but also drawbacks, such as their incapacity to always be systematic and reproducible or the lack of full control. E-textiles have recently made inroads into music performance settings. The work reported in [126] presents a system that allows musicians to manipulate sounds through gestural interactions captured by textile wearable sensors. The sensors embedded in the e-textiles are used to control, in real-time, audio processing algorithms working with content interactively downloaded from the Internet thanks to direct wireless connectivity over a 4G network.

H. SMART MUSICAL INSTRUMENTS AND LOW-LATENCY EMBEDDED AUDIO SYSTEMS

Smart musical instruments are a family of musical instruments recently proposed by Turchet *et al.* [143]. They are characterized not only by sensor interfaces typically used in the so-called *augmented instruments* [99], but also by embedded computational intelligence, a sound processing and synthesis engine, wireless connectivity to local networks or the Internet, an embedded sound delivery system, and an onboard system for feedback to the player. Smart Instruments are capable of directly exchanging musically-relevant information with one another and communicating with a diverse network of external devices (such as smartphones, wearables, virtual reality headsets, or stage equipment).

To our knowledge, the first exemplar of this family of musical instruments is the Sensus Smart Guitar developed by MIND Music Labs [143]. This instrument consists of a conventional electro-acoustic guitar augmented with IoT technologies. Several sensors embedded in various parts of the instrument enable the tracking of a variety of gestures of the musician. Thanks to a low-latency sound engine, the detected gestures are used to modulate sounds produced by the instrument's strings as well as to control the generation of other sounds (e.g., the ones of synthesizers). The sonic output, digitally processed or generated, is delivered by an actuation system applied to the instrument's resonating wooden body. A wireless communication system enables the transmission and reception of different types of data from the instrument to a variety of smart devices and vice versa. Various applications leveraging Sensus's features have been explored, which constitute examples of the novel interconnection and interaction possibilities offered by Smart Instruments within the IoMusT [139]. In more detail, Sensus has been used in the following cases: 1) to wirelessly control external equipment such as projected visuals, stage lights, and digital workstations running on external computers; 2) to simultaneously play together with multiple apps running on connected smartphones and tablets, which allow for the wireless real-time streaming of audio content and/or musical messages interpreted and rendered by Sensus' sound engine; 3) to control elements of VR scenarios provided by VR headsets; 4) to play over music streamed from the cloud and share recordings on social networks.

Another example of a smart instrument is the Smart Cajón prototype reported in [141] and [142], which consists of a conventional acoustic cajón smartified with sensors, Wi-Fi connectivity, motors for vibro-tactile feedback placed inside an attached cushion, and embedded audio and sensors processing. The embedded intelligence of the instrument has been exploited to detect various positions of the player's hits on the instrument using sensor fusion [57], semantic audio [127], and machine learning [24], [49] techniques, along with the identification of the gesture that produced the hit, for repurposing this information into sound and automatic score transcription [15]. The instrument can interact wirelessly with connected external equipment, such as smartphones. Smartphone apps were developed to allow the player to configure the instrument, to provide other musicians or even audience members with interactive control of the instrument's internal status, as well as to deliver information to the player via haptic notifications through the motors embedded in a cushion. The Smart Cajón has been employed in conjunction with musical haptic wearables, where the temporal and intensity information of the hits produced by the player were mapped, in real-time, to tactile stimuli delivered to audience members [137].

The development of a smart instrument is deeply rooted in the embedded platform utilized for all aspects of the embedded intelligence, including low-latency audio processing, sensor processing, and wireless connectivity. To date, most professional audio devices targeting the hard requirements of real-time performance are built using dedicated digital signal processors and ad-hoc real-time operating systems. These are complex to program, offer very limited support to interface to other hardware peripherals, and lack modern software libraries for networking and access to cloud services. This makes the realization of connected products very difficult. The state-of-the-art technology for creating smart musical devices, such as smart instruments, is represented by ELK, an IoT operating system devised for musical applications that is developed by MIND Music Labs[9] and has been released in 2018. ELK is based on Linux, guarantees round-trip latencies of 1 ms, supports music software plugins, offers a range of connectivity options and provides developers with efficient development tools. ELK is embedded in the Sensus Smart Guitar and in other smart devices. An alternative for the maker culture, is Bela, a rapid prototyping board for audio and sensor processing based on the Beaglebone Black platform and open source software [93]. Bela has been utilized to prototype the Smart Cajón described in [141].

[9] www.mindmusiclabs.com/ELK

I. VIRTUAL/AUGMENTED REALITY AND CINEMATIC EXPERIENCES

Virtual and Augmented Reality (VR/AR) interfaces such as Head-Mounted Displays (HMDs) have the potential to act as Musical Things letting users experience or interact with musical content in novel ways. VR and AR offer the opportunity to explore radically new spaces for multisensory musical experiences, as well as for collaborative music making such as telepresence for networked performances, rehearsals, improvisations, compositions, or music teaching.

New forms of musical interactions have been proposed using immersive virtual environments. A number of "virtual reality musical instruments" have been developed, which allow a user to play music within the virtual world (for a recent review see [121]). Immersive music performance systems have been created using 3D reactive widgets, graphical elements that enable efficient and simultaneous control and visualization of musical processes, along with Piivert, an input device developed to manipulate such widgets, and several techniques for 3D musical interaction [17], [18]. Virtual reality systems for collaborative music making are also starting to emerge. A recent example is LeMo a networked virtual reality system that provides two people with a shared musical interface based on a step sequencer, with which they can co-create 8-beat music loops [96]. In a different vein, networked interactions between physical musical instruments and virtual environments for audience members are also starting to be explored. For instance, the Sensus Smart Guitar has been used to wirelessly control elements of virtual environments provided via head-mounted-displays [139].

Augmented Reality (AR) has been used to enhance live concert performance experiences, as well as for participatory performance applications. Mazzanti *et al.* [92] proposed the Augmented Stage, an interactive space for both musicians and audience members, where AR techniques are used to superimpose a performance stage with a virtual environment, populated with interactive elements. Spectators contribute to the visual and sonic outcome of the performance by manipulating virtual objects via their mobile phones. Berthaut *et al.* [19] developed Reflets, a mixed-reality environment that allows one to display virtual content on stage, such as 3D virtual musical interfaces or visual augmentations of instruments and musicians. Poupyrev *et al.* [106] proposed the Augmented Groove, a musical interface for collaborative jamming where AR, 3D interfaces, as well as physical, tangible interaction are used for conducting multimedia musical performance.

The growing availability of 360° videos has recently opened new opportunities for the entertainment industry, so that musical content can be delivered through VR devices. Recent examples include Orchestra VR, a 360° 3D performance featuring the opening of Beethoven's Fifth Symphony performed by the Los Angeles Philharmonic Orchestra, accessible via an app for various VR headsets,[10] Paul McCartney's 360 cinematic concert experience app allowing the experience of recorded concerts with 360° video and 3D audio using Google's Cardboard HMD, Björk's 360° Stonemilker video released as a virtual reality app,[11] and Los Angeles Radio Station KCRW, which launched a VR App for "intimate and immersive musical performances".[12]

IV. EXAMPLES OF IOMUST SCENARIOS

Thanks to the IoMusT it is possible to reimagine certain musical activities such as live music performance or music learning, by leveraging a technological ecosystem that multiplies possibilities of interaction between different actors (e.g., audiences, musicians, audio engineers, students, teachers) through dedicated Musical Things and services, as well as by augmenting the sonic content with other modalities (e.g., vision, touch). This has the potential to revolutionize the way music is composed, learned, experiences, as well as recorded.

What would be the experience of musicians and audience members in settings with the IoMusT? To give an idea of such settings, we sketch five hypothetical scenarios below. These scenarios do not aim at representing real user needs or desires, which need to be investigated. Their role is to enable a discussion about the potential of the IoMusT and to identify capabilities that are currently missing.

Scenario 1 (Augmented and Immersive Concert Experiences): David, Laura, and Vinay, attend a concert of their preferred band and upon arrival they can choose various types of concert experience interfaces. David choses a pair of smart glasses (wearable computer glasses that add information alongside to what the wearer sees), Laura picks up an armband that responds to the music through physical stimulations, and James selects a hoody with sensors and loudspeakers. All these devices can track the movements of the user and transmit this information to connected equipment. While playing, the band musicians act on the sensor interface of their smart instruments to deliver to the audience visuals to be displayed on the smart glasses (dynamic visuals appear in a virtual environment designed to represent the visual identity and the narrative of the music played by the band), vibrations and thermal sensations provided by the wearables' actuators (such as vibration patterns associated to a musician's gesture), as well as complementary sounds produced by the loudspeakers embedded in the smart hoody (such as synthesized sounds synchronized to those produced by a smart bass). In other parts of the concert the smart instruments automatically interpret performers' expression and map associated representations to other modalities: David's smart glasses change the colors of elements from the stage according to the sounds from the smart keyboard; Laura can "feel" the vibrato produced by the smart violin performer and the rhythm produced by the smart drums performer; Vinay can hear a version of the solo of the smart guitar processed with additional audio effects. When David, Laura, and Vinay, and the rest of the audience start to dance, the sensors embedded in their smart devices track their movements and this information is mapped to trigger smoke machines and stage lights.

Scenario 2 (Co-Located Audience Sensing and Remote Audience Participation): Asha goes to a pop music concert and brings the smart wristbands that she normally uses at the gym to monitor her activity. Such a wearable is equipped with sensors capable of tracking Asha's movements and physiological parameters (e.g., heart rate). During the concert the smart wristband interfaces with a system that collects, process and interpret data sensed by the wristbands from Asha and other audience members. The system predicts the audience's mood which is used in different ways: i) to help musicians decide on the next songs in their playlist, ii) by choreographers to create real-time live visualizations responding the mood of the audience, (iii) by the audience itself to control some of the performance experience attributes such as lighting effects (e.g., lights following the mood of the audience). Jié could not attend the live concert because he lives in a city far away from the concert venue. He is however experiencing the concert from his home with his VR device. The VR headset runs a software based on 360° video technologies that allows Jié to virtually experience the show, even to walk on stage and see every detail of the performance of his preferred musician in the band. The VR headset is also equipped with a system capable of tracking Jié's hands. The concert is structured in such a way to involve, at specific times, the remote audience in the music creation process. The specific moments are visually signaled via the display of the VR headset and at those moments Jié, by interacting with some virtual objects manipulated with his hands, can generate sounds that are delivered to the concert venue and diffused via the loudspeakers as accompaniment to the band.

Scenario 3 (Remote Rehearsals and Smart Instruments Preset Sharing): Jennifer and Bob play respectively a smart guitar and a smart double bass. They live in different cities located at 200 Km from each other and need to rehearse for a gig. It is not convenient for them to meet in person, considering the cost of traveling and the time taken. Therefore, they decide to rehearse remotely using a screen-based interface on their laptop from their respective homes and point-to-point audio streaming between their smart instruments, where the sound of Jennifer's smart guitar is received on and reproduced by Bob's smart bass and vice versa. To practice they also use a shared virtual metronome, which is synchronized on both the instruments. At some point, they decide to change the timbral configuration of their instrument for a specific piece. So they

[10] www.laphil.com/orchestravr
[11] www.youtube.com/watch?v=gQEyezu7G20
[12] www.kcrw.com/vr

use their laptops to wirelessly instruct their smart instruments to download from a special social network for musicians a set of sound and audio effect presets for their respective instruments. They then modify their downloaded presets by adjusting parameters using a graphical user interface on their laptops. After this modification they save the parameter configuration as a new preset and finally they upload it onto the cloud so that other members of the social network can download it at a later stage.

Scenario 4 (Music e-Learning): Tanya has just started to self-teach herself how to play a smart ukulele. She is following a method that comes with a game-based app for her smartphone and that is capable of interacting with Amazon Echo's Alexa voice-based assistant. The smart ukulele receives from the app the score that must be followed during the learning game, analyzes in real-time what Tanya plays, infers the errors that she is making, and sends the analysis back to the app and to Alexa. The app displays those errors and Alexa suggests what she should do to avoid making mistakes. Michael, a friend of Tanya, also wants to self-learn how to play a smart ukulele. The program he is following uses a smartphone app similar to the one Tanya is using as well as a pair of smart glasses, both of which wirelessly communicate with the smart ukulele. The smart glasses are equipped with a small video camera and can superimpose visual content to the field of view. When Michael makes an error, the smart glasses display the hands of an avatar that show the correct hand and fingering positions. After their practice sessions, both Tanya and Michael save in a cloud repository the results of their performance. After one week of daily practice, the cloud repository services automatically send a notification to Tanya's and Michael's smartphones recommending what learning program they should follow the week after. These services are overseen by educationalists and instrument teachers, and also rely on music performance data analytics conducted on smart ukuleles from learners of various levels around the world.

Scenario 5 (Smart Studio Production): Andrew is a studio producer. He has audio- and video-recorded a live session of a band and he is producing the mix and the mastering as well as a 360° video. The smart mixing console he used during the live recording has recorded several information streams in addition to the audio signal generated by each of the band's smart instruments: the signals coming from sensors, the metadata associated with the instrument's configurations (e.g., the utilized synthesizers and sound effects as well as their presets), the information about musicians' performative gestures, and the score automatically transcribed by each smart instrument. Thanks to all this information and to machine learning techniques, he can more easily carry out his production. Indeed he has access to the song structure automatically synchronized with the audio track, a tempo track following the temporal variations from the performers, notifications of potential asynchronies between instrumental parts in certain segments highlighted in the digital audio workstation, as well as virtual instrument tracks that were set up with the presets used by the musicians, which he can alter by acting on the corresponding parameters. Then he saves the results of his production in a file encoded according to the "IoMusT studio producer interchange format" and sends it to his studio producer friend Perrine for feedback. Perrine opens the file with a digital audio workstation different from the one Andrew is using and modifies the track of the smart bass adding an equalization to the effects controlled by sensor interface. She then sends back to Andrew the resulting file which contains her suggestions for improvement. Andrew uses the file for the audio of the video of the band. The resulting file is encoded in the "IoMusT media file format" and uploaded on a cloud based repository where the band can download it for review before release.

MISSING CAPABILITIES

The described scenarios embody many key ideas from the IoMusT. Scenario 1 would be possible with Musical Things designed to support novel multisensory experiences augmenting traditional concert situations. The scenario also illustrates the wireless interconnection of different Musical Things and how it is possible to leverage the intelligence embedded in them to create novel forms of interactions between musicians and audience members. These interactions can be supported by radically novel musical instruments capable of controlling networked devices interactively based on user inputs, intelligent sensing, and automatic processes. The examples also involved novel types of wearables (e.g., using e-textiles) that would provide sensory feedback to users and also sense and interpret behavioral information from the audience with the end goal of controlling aspects of the performance.

Scenario 2 relies on affective computing methods to analyze in real-time the physical/physiological responses of users and interpret the mood of the audience. The scenario also provides an example of how co-located and remote audience interventions can be intertwined. The scenario also illustrates how data and services from the IoMusT could be used in virtual/mixed reality environments to create immersive experiences for co-located or remote audiences and ways to collaborate in the music making process. Such scenario highlights the need for Tactile Internet technologies enabling ultra-low latency within the IoMusT. This also calls for novel artistic forms integrating multi-user interactions and new media in meaningful ways. The scenario also indicates the need for interoperable IoT devices that even if designed for other primary applications (e.g., health monitoring wristband) could be repurposed for musical activities.

Scenario 3 shows possibilities enabled by the IoMusT and Tactile Internet to support remote musical experiences that are close to the co-located ones in terms of realistic interactions between musicians. These interactions are also supported by the direct interconnection of interoperable musical instruments, as well as by synchronization mechanisms across a WAN. Moreover, scenario 3 illustrates the use of cloud-based services for musicians that can directly interface with smart instruments.

Scenario 4 shows how techniques for retrieving information from the musician's performance can be exploited to understand musical mistakes and provide feedback to the user. It also illustrates the benefit of having interoperable IoMusT and IoT devices. Scenario 4 also shows the importance of context-awareness and proactivity: the cloud services understand the needs of the two self-taught musicians and provide recommendations combining big data and artificial intelligence.

Scenario 5 shows how recording and studio production procedures could be impacted by smart instruments, thanks to their capacity to provide a set of synchronous contextual information supplementing the audio signal. It also highlights the benefit of having such information encoded in novel interoperable formats that can be used by different programs.

These scenarios may look like science fiction rather than reality today. The reason lies in the several technical, artistic, and pedagogical challenges that these scenarios imply. We discuss some of these challenges in Section V.

V. CURRENT CHALLENGES

IoMusT inherits all the challenges of the general field of IoT (see e.g., [144]). In addition to these, the practical realization of the envisioned IoMusT poses specific technological, artistic, pedagogical, legal, personal data- and creative data-related challenges. In this section, we present an overview of the main challenges for each of these categories.

A. TECHNOLOGICAL CHALLENGES

The realization of the IoMusT vision described in Section II comes about through the evolution of the network and services' infrastructure as well as of the capabilities of Musical Things connecting to them. The connectivity features of the envisioned Musical Things need to transcend beyond the state-of-the-art technologies available today for the music domain. Between a sensor that acquires a measurement of a specific auditory, physiological, or gestural phenomenon, and the receiver that reacts to that reading over a network, there is a chain of networking and information processing components, which must be appropriately addressed in order to enable acceptable musical interactions over the network. Current NMP systems suffer from transmission issues of latency, jitter, synchronization, and audio quality. These hinder real-time interactions that are essential to collaborative music creation [113]. It is also important to notice that for optimal NMP experiences, several aspects of musical interactions must be taken into account beside the efficient transmission of audio content. Indeed, during co-located musical interactions musicians rely on several modalities in addition to the sounds generated by their instruments, which include for instance the visual feedback from gestures of other musicians, related tactile sensations, or the sound reverberation of the space [159]. Providing realistic performance conditions over a network represents a significant engineering challenge due to the extremely strict requirements in terms of network latency and multimodal content quality that are required to achieve a high-quality interaction experience [128]. We identify four key areas that currently maintain obstructions to many interesting IoMusT application scenarios: i) Low-latency, high-reliability, and synchronization; ii) Interoperability and standardization; iii) Musical Things design; and iv) Representation and analysis of multimodal content.

1) LOW-LATENCY, HIGH-RELIABILITY, AND SYNCHRONIZATION

As we have reviewed in Section III-B, one of the most demanding engineering challenges is the transmission of low-latency high quality audio streams over networks, both wireless (especially cellular and local area networks) and wired. The significant engineering challenge to design communication networks capable of supporting true real-time music services is due to the extremely strict requirements in terms of network delay and transmission reliability. These communication requirements are determined by a high-quality interaction experience. Stable message reception rate and a satisfying synchronization between musicians are examples of requirements for high quality interaction over networks. We analyze these two aspects in the following.

The message transmission over a wireless or wired network is always subject to some form of randomness, due to random interference in wireless channels, or random background traffic on wired networks. The result is that the messages have a non-zero probability of not being received. Even when they are received, the reception delay may vary remarkably as a consequence of the randomness described above. The delay in receiving a message can be described by a random variable. The expectation of such a random variable is what is usually called latency, while the standard deviation of such a random variable is defined as jitter. For musical performances, the latency has to be on the order of milliseconds and it is also important to have small jitter, on the order of few milliseconds, because this allows the receiver to adapt to the delay, and in some cases compensate for it [113]. But if the jitter is too large, no such compensations are possible. The jitter is ultimately a measure of the delays that will occur with very high probability within a boundary around the mean. In case of network packet loss, or to compensate for the effects of network jitter, if a packet reaches its destination after its scheduled playback time, its audio data is no longer valid.

Another challenge in music communication services is the synchronization of audio streams produced by devices that do not share the same clock. Even if devices of different networks would initially share the same clock, they need a re-synchronization procedure from time to time. Several protocols have been proposed to achieve such a synchronization [113]. However, existing methods are insufficient as they don't enable to reach low latency requirements. Consequently, a significant ongoing effort is to perform a so-called ''physical layer synchronization'' already at the wireless communication interface. There, the current proposal is to shape the communication waveform so that both music information and synchronization information can be piggybacked on the same waveforms [40].

To overcome the technical problems mentioned above, there have been some attempts that can be generically divided into two categories: design new wireless/wired communication protocols from scratch [86], or optimize existing protocols [101]. The first approach determines a total re-design of wireless protocols such as Wi-Fi or Bluetooth, and therefore there are concerns around the capacity for widespread usage. This approach has the concrete potential to substantially overcome the jitter and synchronization issues mentioned above. However, it would require the adoption of specialized radio chips. The approach is not yet part of standards nor is there an easy way to make it a general commercial standard ahead of other popular standards such as Wi-Fi, Bluetooth, or 5G. There are grounded concerns about the willingness of the industry to consider this method [86]. Therefore, the second approach is more appealing. Such an approach consists of adapting the communication protocol parameters that are free so as to optimize the transmission of music. This has been used to support Musical Instrument Digital Interface (MIDI) over Bluetooth [14], or to optimize the protocol parameters of Wi-Fi for musical performance [101]. However, these attempts are still not satisfactory, mostly due to the inherent limitation of the physical layer of the communication protocol stack and the protocol overhead. How to overcome these issues is currently under investigation. The envisioned Tactile Internet [1], [48] is expected to solve these issues by providing communication networks, both wireless and wired, capable of ensuring ultra-low latency communications, with end-to-end delays on the order or few milliseconds. Along the same lines, the use of edge computing technologies [22], [62], [90], [119], [122], [162] are expected to play a relevant role in the reduction of the latency and bandwidth pressure by offloading the computation from the cloud.

2) INTEROPERABILITY AND STANDARDIZATION

Standardization activities represent a central pillar for the IoMusT realization as the success of IoMusT depends strongly on them. Indeed, standardization provides interoperability, compatibility, reliability, and effective operations on both local and global scales. However, much of this work remains unrealized. More standardized formats, protocols and interfaces need to be built in the IoMusT to provide more interoperable systems. This issue is also common to the more general field of IoT [125]. Within the IoMusT, different types of devices targeting musicians or audiences (both co-located and remote) are used to generate, track, and/or interpret multimodal musical content, and need to be able to dynamically discover and spontaneously interact with heterogeneous computing, physical resources, as well as digital data. Their interconnection poses specific challenges, which include the need for ad-hoc protocols and interchange formats for musically relevant information that have to be common to the different Musical Things, as well as the definition of common APIs specifically designed for IoMusT applications.

To date, musical messages such as OSC and MIDI can be transmitted over a WLAN leveraging standard protocols such as Wi-Fi and Bluetooth that are commonly provided by a variety of smart devices. However, the wireless transmission of audio signals in a low-latency and high quality fashion today relies on bespoke proprietary formats and systems developed by different manufacturers (e.g., wireless transmitter/transceiver for guitars by Line6). No methods are available today to accomplish the low-latency and high quality communication of audio signals over the most widespread standard protocols for wireless communication.

OSC is promising to be part of an IoMusT standard as far as control message interoperability is concerned. However, some aspects of OSC prevent it becoming a standard capable of replacing the widely adopted MIDI standard, which was defined by the musical industry in the early 1980s. For instance, as opposed to MIDI, OSC only describes a communications protocol, not a digital interface and electrical connectors (wired or wireless) that connect devices. Moreover, there is no standard namespace in OSC for interfacing devices. As a result, connected devices do not know of each other nor of each other's capabilities. In addition, a file format for OSC similar to Standard MIDI File is missing, which could be used to share data between different applications. Nevertheless, there are on-going efforts to standardize OSC within the developers community.[13]

On the other hand, MIDI is not well suited to achieving interoperability across heterogeneous devices since it is less flexible, given the fact that it was specifically conceived for communication across musical instruments. However, there are ongoing efforts within the MIDI Manufacturers Association (which gathers the world-leading music hardware and software manufacturers) to define the MIDI HD standard.[14] This is expected to provide the current MIDI standard with new features that have the potential to improve the interoperability aspect.

Besides interoperability, an IoMusT standard should take into account synchronization aspects based on such interoperability. At present, Ableton Link appears to be the ideal candidate for this task as far as LANs and WLANs are concerned. However, currently there is not a corresponding counterpart for WANs [113].

In a different vein, new formats for interchange of files within the IoMusT are also needed. A current standardization effort towards this direction is represented by the *MPEG-A: Interactive Music Application Format (IM AF)* [54], [64], which was developed under the auspices of the International Organization for Standardization/International Electrotechnical Commission Moving Picture Experts Group (MPEG). IM AF combines multiple audio tracks and appropriate additional information, which enables users to experience various preset mixes and to make their own mixes complying with

[13] https://github.com/fabb/SynOSCopy/wiki

[14] https://www.midi.org/articles/midi-manufacturers-investigate-hd-protocol

interactivity rules imposed by the music composers with the aim of fitting their artistic creation. This is expected to enable interoperability among the new interactive music services that have emerged in recent years, which typically use proprietary file formats. However, the IoMusT has a multimodal nature, which motivates the definition of standards for formats that account not only for the transmission of audio signals or control messages for musical purposes, but also multimedia content and the associated metadata. Current efforts towards this direction are represented by the ISO Base Media File Format.[15] It is expected that the combination of IM AF with other emerging technologies such as three-dimensional audio/video as well as content-based search and retrieval, will enable novel music applications and services for both audience members and musicians [64].

In [135], Thalmann *et al.* propose a distributed music format called Dynamic Music Object (DYMO). A DYMO is a flexible and modifiable entity that encompasses a bundle of music files, analytical data extracted from the files using music information retrieval techniques, a structural definition relating the audio and the analytical data, and a playback configuration called rendering, which maps controls to parameters. DYMOs rely on semantic web technologies, namely the Web Ontology Language (OWL)[16] and SPARQL[17] [31], which can be used to express queries across diverse linked data sources (see Section V-A.4). DYMO have been employed for context-dependent and adaptive listening experiences on mobile devices using sensor controls (accelerometer, compass, geolocation, etc) [135], or mood intentions from users [11]. The flexibility and networked nature of DYMOs may prove fruitful for the IoMusT ecosystem.

Common software platforms running inside Musical Things are also needed, along with standards regulating them. To date, MIND Music Labs' ELK IoT music operating system represents the first prominent effort towards this direction.

3) MUSICAL THINGS DESIGN

While the design of musical interfaces is a mature subject [65], [99], the setting of an emerging IoMusT ecosystem motivates an expanded view of musical interface design that is rarely tackled today. Keller and Lazzarini articulate this lack as follows: *"Musical activities that take place outside of traditional venues and that feature the audience as an active creative partner demand design techniques that are not currently supported by mainstream musical interaction approaches"* [72].

While Musical Things can be used locally, their distribution and heterogeneity is an important feature of the system. This suggests that one ought to think of interfaces and interactions not just as a topic of local actors acting on local interfaces, but as one or more local or distributed actors acting on local and distributed interfaces. A potential way into thinking along those lines was suggested by Fencott and Bryan-Kinns [47], who proposed that multi-actor musical performances should be studied along the lines of collaborative human-computer interaction research and based on methodologies long established in the field of computer-supported cooperative work. This view of the social, collaborative and interactional study of Internet of Things in general has also been suggested by Robertson and Wagner [109].

The heterogeneous infrastructure of the IoMusT introduces new opportunities for the design of Musical Things, which bears several technical challenges. Firstly, energy consumption aspects should be taken into account in the design of Musical Things to fully leverage the ubiquitous potential of different types of devices (e.g., smart instruments). Secondly, miniaturization of computing units dedicated to low-latency sound processing, sensing, communication, and power supply is required in order to embed such units in musical instruments or wearables and make them light.

A major challenge concerns the design of Musical Things as entities capable of supporting effective interactions with their users [114]. This is an issue present in the more general field of IoT, where the human is often neglected in IoT design [76], [77], [100]. Therefore, research efforts are needed to understand how humans should interact with Musical Things and to define rich interaction paradigms that could enable the users to leverage the IoMusT potentialities and benefits. Possible solutions to this challenge could be found in co-design procedures where Musical Things are designed together with their end users (see e.g., [141]).

Wearable systems present many opportunities for novel forms of musical interaction, especially involving multiple sensory modalities. Related design challenges concern the optimization of the sensor and actuator capabilities of these devices for musical purposes (e.g., temporal precision, low latency, synchronization of audio, visual, and tactile modalities). Another related challenge is how to effectively use multiple sensory modalities in live music performances, including those that involve participation of the audience. In particular the sense of touch leveraged by musical haptic wearables [138] could have a high impact on the musical experience of the audience.

To date, the opportunity of using VR to build completely novel social and cooperative experiences bridging the gap between musicians and audience has not been explored. One of the big challenges currently in VR is how to achieve a seamless interaction with virtual worlds, proliferating many different interfaces [128].

These are but a few examples of the wide range of issues related to the design of Musical Things and how participants experience their use. In fact, the IoMusT provides issues induced from novel and diversified settings, scale and distribution, collaboration and communication, transparency and affordance, access and privacy, and a range of interaction

[15] https://mpeg.chiariglione.org/standards/mpeg-4/iso-base-media-file-format

[16] http://www.w3.org/TR/owl2-overview/

[17] http://www.w3.org/TR/sparql11-query/

types from real-time interactive (such as live performances) to highly asynchronous (such as collaborative authoring of compositions).

4) REPRESENTATION AND ANALYSIS OF MULTIMODAL CONTENT

Signals associated with various sensory modalities (e.g., sound, light, pressure, temperature) may be exchanged within the IoMusT representing phenomena as varied as performers' musical controls [133], audiences' physiological responses [131], or information about the musical environment. Musical Things can be developed to dynamically capture, through their sensors, rich intangible cultural expressions linked to music production or reception. For example, gesture-related information can reveal information about co-expressive elements present in the communication process of emotions [32]. Harnessing such intangible information during musical activities relying on IoMusT could provide ways to study musical practice and audience reception, a topic which has received a growing attention in music psychology [41]. But music-related sensor information can also be used for creative mappings between different modalities (one to one, one to many, many to one). An example of such usage is discussed in [131] which proposes a system to generate music visualizations reacting in real-time to listener's arousal response measured with biosensors. Scaling such systems to large audiences could provide radically novel ways to engage audience in creative participation during live music performances. These types of opportunities enabled by the IoMusT may be explored by developing techniques based on multimodal machine learning and semantic audio.

Multimodal machine learning is a fertile area for development of systems capable of deriving meaning from sensor signals from multiple modalities. The work reported in [8] presents the following core challenges for multimodal machine learning: *representation*: how to represent and summarize multimodal data given complementarity and redundancy of multiple modalities; *translation or mapping*: how to translate data from one modality to another given the heterogeneous nature of data and perceptual relationships across domains; *alignment*: how to identify the direct relations between (sub)elements from two or more different modalities; *fusion*: how to join information from two or more modalities to perform a prediction (e.g., joining video, motion capture kinematic parameters and audio to predict relationships between a conductor's gesture and sound, as in [117]); *co-learning*: how to transfer knowledge between modalities, their representation, and their predictive models. Alignment and mapping/translation techniques could be developed to enable "semantic information integration" [21] within the IoMusT ecosystem. A major challenge for IoMusT applications consists of determining flexible mapping strategies to go from the signal to the perceptual domains that enable meaningful and relevant interactions for both experts and novices. These mappings could be based on features extracted in real-time from sensor data and musical audio analysis. Multi-sensor data fusion techniques [57] could be exploited for this purpose, which explicitly account for the diversity in acquired data (e.g., in relation to sampling rates, dimensionality, range, and origin).

The field of semantic audio [127] has evolved to develop computational models extracting high level meaning from audio signals[18] which can be interpreted by humans and machines alike. The semantic audio endeavor combines music information retrieval [26] and the semantic web [60], [31] to map audio signals to machine-readable data interchange formats such as Resource Description Framework (RDF). By essence, RDF extends the linking structure of the Web with expressions of the form *subject–predicate–object* called triples (a subject denotes a resource and a predicate expresses a relationship between the subject and the object). For example, one way to represent the description "The song is in D dorian" in RDF is to use a triple with "The song" as subject, "is in" as predicate relating to musical mode, and "D dorian", as the object. The types, properties, and relationships between concepts in a particular domain are described by ontological models also called vocabularies or knowledge graphs (see e.g., [107], [127] for music). Using such models, structured and semi-structured data can be mixed, exposed, and shared across different applications using Uniform Resource Identifiers (URIs). Some authors have initiated works to automatically infer ontologies from audio signals, for example for the classification of musical instruments and their properties [75]. Semantic web technologies are promising to address interoperability issues in the context of the IoMusT. This would imply the development of appropriate ontologies aiming at describing the music production and reception processes mediated by Musical Things. Such ontologies would enable retrieval of data within the ecosystem given requests from specific creative music services, but would also make logical inferences (e.g., retrieving, from a smart guitar signal, information about note vibrato to generate haptic feedback on audience haptic wearables).

Semantic audio has mostly concentrated on the analysis of recorded music in an "offline" context and several applications leveraging big music data have been proposed, e.g., for computational musicology [154], the analysis of chord progressions [12], perceived emotions [116], or music discovery [157] and recommendation [10]. However, less attention has been paid to real-time application of semantic audio necessary in the context of live musical interactions. A large number of semantic audio signal processing techniques require large segments of audio to make predictions which prevents real-time applications. The Vamp plugin software framework for audio feature extraction [30] only allows implementation of non causal algorithms. The Essentia C++ audio and music analysis library allows real-time computations but not all algorithms available in the library are suited for real-time analysis due to their

[18] Here we distinguish high level descriptors from the semantic domain (e.g., a chord, an emotion) to low level descriptors from the signal domain.

computational complexity [20]. Future research directed at developing real-time semantic audio would benefit IoMusT applications. A recent initiative enables audio feature extraction within the browser [136], which offers interesting semantic audio applications within a Web of Musical Things. Musical Things, cloud computing, and edge computing technologies may offer pathways to develop real-time semantic audio techniques by leveraging multiple modalities and distributed processing power. Musical things for performers such as smart musical instruments [143] present the advantage of being able to capture input control gestures encompassing musical interpretation and expressive information thanks to the embedded sensors. Conducting music information retrieval using both sounds and sensor signals characterizing the causal phenomena at the source of the generated sound may help overcome the so-called "glass ceiling" effect limiting recognition accuracy [4]. The use of source control signals could also limit effects of the acoustics environment such as reverberation, which have been shown to affect audio-content based automatic recognition of musical instruments [13].

New analytic tools are needed to make the most of the IoMusT. Such tools should be able to process large amounts of music-related data and extract meaningful information given tight temporal constraints. Deep learning [81] offers encouraging ways to obtain high level features that could capture the essence of human expression and phenomena associated to musical activities.

From an industrial perspective, IoMusT has the potential to generate new business models proposing ways to use varied types of information collected by Musical Things in meaningful ways. For instance, such information could be used to understand performer and audience behavior, to deliver specific music services (e.g. for learning or co-creation), to enhance concert experiences and increase active engagement. These would entail identification and interception of appropriate "business moments", as defined by Gartner Inc.[19]

B. ARTISTIC CHALLENGES

The IoMusT differentiates itself from standard questions of the IoT by its artistic and performative application. As previously unveiled by ubimus research [71], the enhanced availability of resources and the lack of relevance of the individualistic views on creativity may pose several challenges to IoMusT scenarios. While the scope of artistic potential is wide, we will here consider a few specific aspects.

1) DISTRIBUTED AND SITUATED PERFORMANCES

The emergence of the Internet has boosted the potential for remote interaction and collaborative music making. Music making involves an intense exchange of sonic resources among the stakeholders of musical experiences. This exchange entails a process of decision making which - given the right conditions - does not need to enforce hierarchical or synchronous interactions [70], [84]. Hence, the traditional models for the social organization of musical performance - such as the orchestra or the band - may well be disrupted as well as potentially enhanced by new potentials of this paradigm. Massive connectivity involves asynchronous forms of interaction and exchange of resources without the need for face-to-face communication. Paradoxically, while the network infrastructure promotes collaboration without demanding physical co-presence, an increased miniaturization of the personal electronic devices affords music making in places that previously were not available for creative group activities. This type of enhanced portability is a necessary feature of technology for music making in everyday settings. Both massive connectivity and enhanced portability foster new forms of creative manifestations by targeting everyday contexts and by providing access to distributed resources. Thus, the IoMusT may provide a fruitful ground for further developments in ubiquitous musical activities.

2) COMPOSING THE NETWORK

Networking of Musical Things provides a new challenge for structured musical creation. Traditionally intentional creation of premeditated music is referred to as composition. As proposed by ubimus initiatives, one can think of new forms of composition in an IoMusT ecosystem. Potentially very large-scale infrastructure of musical entities needs to be organized and controlled for an artistic aim. The source of the performance can be as diverse as the kinds of sonic outcomes. For example, city-wide sensor networks could be used to construct trans-national performances where the state of one city becomes the score for performers in another. A composer will need tools to support and control distributed, heterogeneous capabilities. Once a composition is created there is also need for recall and reproduction. Collaborative composition practices further complicate the problem, suggesting a need to investigate collaborative support that can handle large scale participation.

3) IOMUST FOR AUGMENTED PERFORMANCE

Although the IoMusT lends itself well to participatory performance where frontiers between musicians and audiences are blurred, it can also be used to augment presentational performances in various ways. New artistic narratives and content must be produced to test the aesthetics and immersive capabilities and viability of the IoMusT. Composing for the IoMusT in a multimodal way could involve different creative industry sectors like in movie or game productions, where visual and music narratives are combined. Deciding which information to present to audiences to transform their experiences using Musical Things or which information to sense from them for novel mechanisms of engagement represent design challenges that should be tackled by interdisciplinary teams of artists, researchers and technologists.

[19] www.gartner.com/newsroom/id/2602820

C. PEDAGOGICAL CHALLENGES

The IoMusT promises to expand the landscape of music pedagogy technologies. At the core of learning is conveying relevant information and feedback to lead to understanding and improvement.

1) SMART INSTRUMENTS AND SCORES

One of the biggest issues in today's instrumental music education technologies is the lack of solutions effectively capable of providing useful information about how musicians play. Smart instruments may be a solution to this issue. In a conventional context, human teachers giving a lesson in person can watch students playing, and on the basis of their observations can then provide the students with instantaneous feedback on the errors they are making and suggest how to fix them. To date, the most widespread technological solutions designed for music learning, such as apps like Yousician, make use of the audio signal captured by a smartphone microphone to infer information on the notes played (via real-time music information retrieval techniques) and provide the users with recommendations in a gamified way. However, in this process a lot of information is lost as the app cannot infer the exact playing technique used by the musician or whether he/she is holding the instrument in the right way. Conversely, a smart instrument equipped with embedded gesture tracking and a microphone system, which might even be used in conjunction with external equipment such as a videocamera or smart glasses, could provide music learning apps and services with richer and more useful data on how the musician is playing it, compared to the range of information that can be extracted solely using the microphone of a smartphone.

Therefore, new sensor fusion techniques are needed to merge the information coming from different sources (e.g., embedded sensors and microphones, external equipment) along with methods to infer errors made by the musicians and effective human-computer interaction strategies to provide the most useful recommendations.

2) RESHAPING INDIVIDUAL TEACHING AND LEARNING

Traditionally, music training is intense and requires close tutoring. Classical musical instrument training is dominated by 1-on-1 interaction with a teacher interspersed with prolonged and often repetitive isolated practice by the learner. The IoMusT allows one to enhance and diversify this experience. For example, smart instruments can retain records of performance for study and analysis, allowing a closer look at mistakes and how to fix them. This can even be removed from the student-teacher interaction and used later by both student and teacher to review a problem and find strategies for improvement. IoMusT applications can potentially remove the requirement for students and teachers to be co-located by providing detailed networked information exchange of deep aspects of the performance such as information from the student's and teacher's instruments, the ability to highlight or actuate fingering or performance aspects directly on the instrument for the student, and the ability to communicate through video or via annotations on a live score. Teaching sessions can be stored as data-rich repositories and used for later recall, discussion, or potential use in a different pedagogical interaction.

3) SCALING INSTRUMENTAL MUSIC PEDAGOGY

The networked character of the IoMusT paradigm enables scaling, while also offering technologies that mediate the challenges that come from engaging with larger numbers of participants. Enriched information can help a teacher or ensemble conductor to better understand the components of performance and identify areas that need improvement more easily. Feedback can be technologically supported (such as pointers on interactive musical scores). By offering sensor and data rich outcomes from musical instruments and instructor feedback, a synchronous experience can potentially be scaled up to be used as asynchronous learning material useful to many. An important overarching challenge is the distillation of information to the most pertinent aspects and the support of meaningful interaction on the material.

D. PRIVACY, SECURITY, AND LEGAL CHALLENGES

The advent of digitization and the Internet had an enormous impact on the music industry [130], in particular on its laws and economics [95]. In the same way, the IoMusT paradigm brings challenges related to personal data, since Musical Things have the capability of automatically collecting, analyzing, and exchanging data related to their users. New business models can emerge leveraging IoMusT data, for example to provide services related to musical activities (such as intelligent music production, music recommendation, analysis of audience behavior and engagement). Ethical and responsible innovation are crucial aspects to take into account in the design of such services to ensure they are socially desirable and undertaken in the public interest. In addition, music composition is an activity subjected to intellectual property infringements, which has legal implications also for the IoMusT. The issues related to personal and creative data in the IoMusT are in part common to those of the more general IoT field [94] and in part to those of the music industry [42], [95]. Addressing these issues represent a set of challenges that are summarized as follows:

(i) Security: As Musical Things are wireless devices they are subjected to the security risks of wireless communications. In today's Internet, encryption is a key aspect to ensure information security in the IoT. Therefore, Musical Things should be designed to support robust encryption, which poses the challenge of making these devices powerful enough to support it. On the other hand, to enable encryption on Musical Things, it is necessary to make algorithms more efficient and less energy-consuming, along with the development of efficient key distribution schemes [155]. Importantly, a uniform security standard should be developed by the IoMusT research community and industry in order to ensure the safety of the data collected by Musical Things. This challenge is currently unsolved also in the IoT field [148].

(ii) Privacy: Given the pervasive presence of the IoMusT, transparent privacy mechanisms will have to be implemented on a diverse range of Musical Things as well as on the platforms that support them. It is necessary to address issues of data ownership in order to ensure that Musical Things users feel comfortable when participating in IoMusT-enabled activities. IoMusT users must be assured that their data will not be used without their consent. The nature of the data to be kept private represents an ontological and ethical problem [50], [35]. Concerning the IoT field, Weber recently highlighted the growing need for technical and regulatory actions capable of bridging the gap between the automatic data collection by IoT devices and the rights of their users, who are often unaware of the potential privacy risk to which they are exposed [148]. The same inevitably applies to the IoMusT. The definition of privacy policies is one approach to ensure the privacy of information. Musical Things can be equipped with machine-readable privacy policies, so that when they come into contact they can each check the other's privacy policy for compatibility before communicating [111]. Based on this, it is therefore crucial that Musical Things designers and manufacturers adopt a privacy by design approach as well as incorporate privacy impact assessments into the design stage of Musical Things.

(iii) Legal Issues: The IoMusT will induce new legal challenges that must be addressed and it is plausible to expect the need for a new legal environment specific to the IoMusT. For instance, new legal approaches for the protection of privacy and copyright might need to be developed. The current copyright and intellectual property laws, which enable owners of musical content to control the reproduction, distribution, and public performance of their works, might need to be adapted to future IoMusT scenarios. Digital piracy over the Internet is still a problem in the music industry and it is plausible to expect that this issue will persist within the IoMusT. For the IoT field, Weber suggested that the IoT governance should not be dictated by a single group, but that a broad-based stakeholder approach to governance is necessary [147]. The same suggestion holds for the IoMusT.

Ultimately, key to the success of the IoMusT will be the consumers' confidence. Music hardware and software manufacturers will need to convince consumers that the use of Musical Things is safe and secure and to do this, much work is still needed.

VI. CONCLUSIONS

This paper presented a vision for the emerging research field of Internet of Musical Things, which stems from many lines of existing research including Internet of Things, new interfaces for musical expression, networked music performance systems, ubiquitous music, artificial intelligence, human-computer interaction, and participatory art. The IoMusT relates to wireless networks of smart devices dedicated to musical purposes, which allow for various forms of interconnection among musicians, audio engineers, audiences, and educationalists, in both co-located and remote settings. The IoMusT vision offers many unprecedented opportunities but also poses both technological and non-technological challenges that we expect will be addressed in upcoming years by both academic and industrial research.

In this paper, we have argued that one of the most demanding engineering challenges is the transmission of low-latency high quality audio (and in general multimodal) streams over networks, both wireless and wired. The challenge of designing communication networks capable of supporting true real-time music services needs to be addressed by developing fundamentally new methods for low latency and stable message reception rates. The message transmission over a wireless or wired network is always subject to some forms of randomness, due to the random interference in wireless channels, or the random background traffic on wired networks. This gives random delays between the transmission and the reception of messages that are in general difficult to control. To overcome these random latencies, new methods will have to be investigated at the physical layer (e.g., new modulation formats), at the Medium Access Control layer (e.g., network coding and data transmission rates according to machine learning methods), and routing layer (e.g, low latency path selection optimization). An extension of the fundamental optimization theory methods applied to the communication layers mentioned above will be needed. Optimization theory with fast computational algorithms working in real-time will be essential to optimize the solution of modulation, medium access control, and networking problems. These methods will have to be applied to 5G communication networks as well as to Wireless Local Area Networks so that they will be able to ensure low latency for musical applications.

The success of the IoMusT strongly depends on standardization activities, which are currently unrealized. The definition of standards for formats, protocols, and interfaces will allow for the achievement of interoperability between systems. Issues related to security and privacy of information, which are common to the more general field of Internet of Things, should also be addressed, especially for IoMusT systems deployed for the masses. Moreover, research will need to address the challenge of how to design systems capable of supporting rich interaction paradigms that enable users to fully exploit potentialities and benefits of the IoMusT. Multimodal machine learning and semantic audio are expected to play a big role within such a context.

The IoMusT vision imposes a rethinking of music composition practices, which will need to consider the distributed nature of musicians and audiences within the ecosystem, along with the multimodality of the musical content. In addition, the envisioned smart instruments and their interoperability with a variety of Musical Things have the potential to greatly impact how music is composed, played, recorded, taught, and experienced. A framework such as the IoMusT and what it entails for artistic and pedagogical agendas will need to be analyzed. This could pave the way for novel research on audience reception, interactive arts, education and aesthetics.

REFERENCES

[1] A. Aijaz, M. Dohler, A. H. Aghvami, V. Friderikos, and M. Frodigh, "Realizing the tactile Internet: Haptic communications over next generation 5G cellular networks," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 82–89, Apr. 2017.

[2] M. Armbrust *et al.*, "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[3] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.

[4] J. J. Aucouturier and F. Pachet, "Improving timbre similarity: How high's the sky?" *J. Negative Results Speech Audio Sci.*, vol. 1, no. 1, pp. 1–13, 2004.

[5] M. A. J. Baalman, H. Smoak, C. L. Salter, J. Malloch, and M. M. Wanderley, "Sharing data in collaborative, interactive performances: The senseworld datanetwork," in *Proc. Conf. New Interfaces Musical Expression*, 2009, pp. 131–134.

[6] P. Bahadoran, A. L. Benito, T. Vassallo, and J. D. Reiss, "FXive: A Web platform for procedural sound synthesis," in *Proc. Audio Eng. Soc. Conv.*, 2018, pp. 1–5.

[7] D. Baker and D. Müllensiefen, "Hearing wagner: Physiological responses to Richard Wagner's der ring des nibelungen," in *Proc. Int. Conf. Music Perception Cogn.*, 2014.

[8] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8269806

[9] Á. Barbosa, "Displaced soundscapes: A survey of network systems for music and sonic art creation," *Leonardo Music J.*, vol. 13, pp. 53–59, Dec. 2003.

[10] M. Barthet, A. Anglade, G. Fazekas, S. Kolozali, and R. Macrae, "Music recommendation for music learning: Hotttabs, a multimedia guitar tutor," in *Proc. Workshop Music Rec. Discovery*, 2011, pp. 7–13.

[11] M. Barthet, G. Fazekas, A. Allik, F. S. Thalmann, and M. B. Sandler, "From interactive to adaptive mood-based music listening experiences in social or personal contexts," *J. Audio Eng. Soc.*, vol. 64, no. 9, pp. 673–682, 2016.

[12] M. Barthet, M. D. Plumbley, A. Kachkaev, J. Dykes, D. Wolff, and T. Weyde, "Big chord data extraction and mining," in *Proc. Conf. Interdiscipl. Musicol.*, 2014, pp. 1–7.

[13] M. Barthet and M. B. Sandler, "On the effect of reverberation on musical instrument automatic recognition," in *Proc. Audio Eng. Soc. Conv.*, 2010, pp. 1–8.

[14] P. Bartolomeu, J. A. Fonseca, P. Duarte, P. M. Rodrigues, and L. M. Girão, "Midi over Bluetooth," in *Proc. IEEE Conf. Emerg. Technol. Factory Autom.*, Sep. 2005, pp. 95–102.

[15] E. Benetos, S. Dixon, D. Giannoulis, H. Kirchhoff, and A. Klapuri, "Automatic music transcription: Challenges and future directions," *J. Intell. Inf. Syst.*, vol. 41, no. 3, pp. 407–434, 2013.

[16] B. Bengler and N. Bryan-Kinns, "Designing collaborative musical experiences for broad audiences," in *Proc. ACM Conf. Creativity Cogn.*, 2013, pp. 234–242.

[17] D. Berthaut, M. Desainte-Catherine, and M. Hachet, "Interacting with 3D reactive widgets for musical performance," *J. New Music Res.*, vol. 40, no. 3, pp. 253–263, 2011.

[18] F. Berthaut and M. Hachet, "Spatial interfaces and interactive 3D environments for immersive musical performances," *IEEE Comput. Graph. Appl.*, vol. 36, no. 5, pp. 82–87, Sep. 2016.

[19] F. Berthaut, D. M. Plasencia, M. Hachet, and S. Subramanian, "Reflets: Combining and revealing spaces for musical performances," in *Proc. Conf. New Interfaces Musical Expression*, 2015, pp. 1–6.

[20] D. Bogdanov *et al.*, "Essentia: An audio analysis library for music information retrieval," in *Proc. Int. Soc. Music Inf. Retr. Conf.*, 2013, pp. 493–498.

[21] H. Boley and E. Chang, "Digital ecosystems: Principles and semantics," in *Proc. IEEE Int. Conf. Digit. Ecosyst. Technol.*, Feb. 2007, pp. 398–403.

[22] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. 1st Ed. MCC Workshop Mobile Cloud Comput.*, 2012, pp. 13–16.

[23] E. Borgia, "The Internet of Things vision: Key features, applications and open issues," *Comput. Commun.*, vol. 54, pp. 1–31, Dec. 2014.

[24] W. Brent, "A timbre analysis and classification toolkit for pure data," in *Proc. Int. Comput. Music Conf.*, 2010, pp. 1–6.

[25] M. Buffa and J. Lebrun, "Real time tube guitar amplifier simulation using WebAudio," in *Proc. Web Audio Conf.*, 2017, pp. 1–9.

[26] J. A. Burgoyne, I. Fujinaga, and J. S. Downie, "Music information retrieval," in *A New Companion to Digital Humanities*. Hoboken, NJ, USA: Wiley, 2016, pp. 213–228. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118680605.ch15

[27] J. P. Cáceres and C. Chafe, "JackTrip: Under the hood of an engine for network audio," *J. New Music Res.*, vol. 39, no. 3, pp. 183–187, 2010.

[28] J. P. Cáceres, R. Hamilton, D. Iyer, C. Chafe, and G. Wang, "To the edge with China: Explorations in network performance," in *Proc. Int. Conf. Digit. Arts*, 2008, pp. 61–66.

[29] J.-P. Cáceres and A. B. Renaud, "Playing the network: The use of time delays as musical devices," in *Proc. Int. Comput. Music Conf.*, 2008, pp. 1–6.

[30] C. Cannam, C. Landone, and M. Sandler, "Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files," in *Proc. ACM Multimedia Int. Conf.*, 2010, pp. 1467–1468.

[31] C. Cannam, M. Sandler, M. O. Jewell, C. Rhodes, and M. d'Inverno, "Linked data and you: Bringing music research software into the semantic Web," *J. New Music Res.*, vol. 39, no. 4, pp. 313–325, 2010.

[32] B. Caramiaux, F. Bevilacqua, and N. Schnell, "Towards a gesture-sound cross-modal analysis," in *Proc. Int. Conf. Gesture Embodied Commun. Hum.-Comput. Interact.* Berlin, Germany: Springer-Verlag, 2010, pp. 158–170.

[33] A. D. de Carvalho, Jr., M. Queiroz, and G. Essl, "Computer music through the cloud: Evaluating a cloud service for collaborative computer music applications," in *Proc. Int. Comput. Music Conf.*, 2015, pp. 226–233.

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

[35] S. Chen and M. A. Williams, "Privacy: An ontological problem," in *Proc. Pacific Asia Conf. Inf. Syst.*, 2010, pp. 1402–1413.

[36] A. Clément, F. Ribeiro, R. Rodrigues, and R. Penha, "Bridging the gap between performers and the audience using networked smartphones: The *a.bel* system," in *Proc. Int. Conf. Live Interfaces*, 2016.

[37] *Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANS)*, IEEE Standard 802.15.4, LAN/MAN Standards Committee, 2003.

[38] W. Dargie and C. Poellabauer, *Fundamentals of Wireless Sensor Networks: Theory and Practice*. Hoboken, NJ, USA: Wiley, 2010.

[39] A. Deusany, S. W. Lee, and G. Essl, "Understanding cloud service in the audience participation music performance of *crowd in c [loud]*," in *Proc. Conf. New Interfaces Musical Expression*, 2016, pp. 1–6.

[40] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[41] M. C. Dobson and J. Sloboda, "Staying behind: Explorations in post-performance musician–audience dialogue," in *Coughing and Clapping: Investigating Audience Experience*, vol. 1. Farnham, U.K.: Ashgate, 2014, pp. 1–17.

[42] R. F. Easley, "Ethical issues in the music industry response to innovation and piracy," *J. Bus. Ethics*, vol. 62, no. 2, pp. 163–168, 2005.

[43] G. Essl, "Automated ad hoc networking for mobile and hybrid music performance," in *Proc. Int. Comput. Music Conf.*, 2011, pp. 1–4.

[44] G. Essl and S. W. Lee, "Mobile devices as musical instruments—State of the art and future prospects," in *Proc. Int. Symp. Comput. Music Multidiscipl. Res.*, 2017, pp. 1–12.

[45] G. Essl, M. Rohs, and S. Kratz, "Use the force (or something)—Pressure and pressure-like input for mobile music performance," in *Proc. Conf. New Interfaces Musical Expression*, 2010, pp. 1–4.

[46] G. Fazekas, M. Barthet, and M. B. Sandler, "Novel methods in facilitating audience and performer interaction using the mood conductor framework," in *Sound, Music, and Motion*, M. Aramaki, O. Derrien, R. Kronland-Martinet, and S. Ystad, Eds. Cham, Switzerland: Springer, 2014, pp. 122–147.

[47] R. Fencott and N. Bryan-Kinns, "Computer musicking: HCI, CSCW and collaborative digital musical interaction," in *Music and Human-Computer Interaction*. London, U.K.: Springer, 2013, pp. 189–205.

[48] G. P. Fettweis, "The tactile Internet: Applications and challenges," *IEEE Veh. Technol. Mag.*, vol. 9, no. 1, pp. 64–70, Mar. 2014.

[49] R. Fiebrink and B. Caramiaux, "The machine learning algorithm as creative musical tool," in *Oxford Handbook of Algorithmic Music*, R. Dean and A. McLean, Eds. London, U.K.: Oxford Univ. Press, 2016.

[50] L. Floridi, "The ontological interpretation of informational privacy," *Ethics Inf. Technol.*, vol. 7, no. 4, pp. 185–200, 2005.

[51] A. Fraietta, "Open sound control: Constraints and limitations," in *Proc. Conf. New Interfaces Musical Expression*, 2008, pp. 19–23.

[52] J. Freeman, "Large audience participation, technology, and orchestral performance," in *Proc. Int. Comput. Music Conf.*, 2005, pp. 1–4.
[53] L. Gabrielli and S. Squartini, *Wireless Networked Music Performance*. Singapore: Springer, 2016.
[54] J. C. García, P. Kudumakis, I. Barbancho, L. J. Tardón, and M. Sandler, "Enabling interactive and interoperable semantic music applications," in *Springer Handbook of Systematic Musicology*. Berlin, Germany: Springer, 2018, pp. 911–921.
[55] D. Guinard, V. Trifa, F. Mattern, and E. Wilde, "From the Internet of Things to the Web of things: Resource-oriented architecture and best practices," in *Architecting the Internet of Things*. Berlin, Germany: Springer, 2011, pp. 97–129.
[56] K. Ha *et al.*, "The impact of mobile multimedia applications on data center consolidation," in *Proc. IEEE Int. Conf. Cloud Eng.*, Mar. 2013, pp. 166–176.
[57] D. L. Hall and J. Llinas, *Multisensor Data Fusion*. Boca Raton, FL, USA: CRC Press, 2001.
[58] A. Hazzard, S. Benford, A. Chamberlain, C. Greenhalgh, and H. Kwon, "Musical intersections across the digital and physical," in *Proc. Digit. Music Res. Netw. Abstr.*, 2014, pp. 1–2.
[59] A. Hindle, "CloudOrch: A portable soundcard in the cloud," in *Proc. Conf. New Interfaces Musical Expression*, 2014, pp. 277–280.
[60] P. Hitzler, M. Krotzsch, and S. Rudolph, *Foundations of Semantic Web Technologies*. Boca Raton, FL, USA: CRC Press, 2009.
[61] O. Hödl, G. Fitzpatrick, F. Kayali, and S. Holland, "Design implications for technology-mediated audience participation in live music," in *Proc. Sound Music Comput. Conf.*, 2017, pp. 28–34.
[62] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—A key technology towards 5G," ETSI, Sophia Antipolis, France, ETSI White Paper 11, 2015, pp. 1–16.
[63] P. Ignoto, I. Hattwick, and M. M. Wanderley, "Development of a vibrotactile metronome to assist in conducting contemporary classical music," in *Proc. Int. Conf. Appl. Hum. Factors Ergonom.*, 2017, pp. 248–258.
[64] I. Jang, P. Kudumakis, M. B. Sandler, and K. Kang, "The MPEG interactive music application format standard [standards in a nutshell]," *IEEE Signal Process. Mag.*, vol. 28, no. 1, pp. 150–154, Jan. 2011.
[65] A. R. Jensenius and M. J. Lyons, Eds., *A NIME Reader: Fifteen Years of New Interfaces for Musical Expression*. Cham, Switzerland: Springer, 2017.
[66] X. Jiang *et al.*, "Low-latency networking: Where latency lurks and how to tame it," *Proc. IEEE*, to be published. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8452158
[67] P. Johansson, M. Kazantzidis, R. Kapoor, and M. Gerla, "Bluetooth: An enabler for personal area networking," *IEEE Netw.*, vol. 15, no. 5, pp. 28–37, Sep. 2001.
[68] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner, "The reactable: Exploring the synergy between live music performance and tabletop tangible interfaces," in *Proc. 1st Int. Conf. Tangible Embedded Interaction*, 2007, pp. 139–146.
[69] F. Kayali *et al.*, "Playful technology-mediated audience participation in a live music event," in *Proc. Extended Abstr. Publ. Annu. Symp. Comput.-Hum. Interact. Play*, 2017, pp. 437–443.
[70] D. Keller, "Compositional processes from an ecological perspective," *Leonardo Music J.*, vol. 10, pp. 55–60, Dec. 2000.
[71] D. Keller, L. V. Flores, M. S. Pimenta, A. Capasso, and P. Tinajero, "Convergent trends toward ubiquitous music," *J. New Music Res.*, vol. 40, no. 3, pp. 265–276, 2011.
[72] D. Keller and V. Lazzarini, "Ecologically grounded creative practices in ubiquitous music," *Org. Sound*, vol. 22, no. 1, pp. 61–72, 2017.
[73] D. Keller, V. Lazzarini, and M. S. Pimenta, *Ubiquitous Music*. Cham, Switzerland: Springer, 2014.
[74] D. Keller *et al.*, "Ecologically grounded multimodal design: The Palafito 1.0 study," in *Proc. Int. Comput. Music Conf.*, 2014, pp. 1–8.
[75] S. Kolozali, M. Barthet, G. Fazekas, and M. Sandler, "Automatic ontology generation for musical instruments based on audio analysis," *IEEE Trans. Audio, Speech, Language Process.*, vol. 21, no. 10, pp. 2207–2220, Oct. 2013.
[76] T. L. Koreshoff, T. W. Leong, and T. Robertson, "Approaching a human-centred Internet of Things," in *Proc. Austral. Comput.-Hum. Interact. Conf.*, 2013, pp. 363–366.
[77] T. L. Koreshoff, T. Robertson, and T. W. Leong, "Internet of Things: A review of literature and products," in *Proc. Austral. Comput.-Hum. Interact. Conf.*, 2013, pp. 335–344.
[78] J. P. Lambert, S. Robaszkiewicz, and N. Schnell, "Synchronisation for distributed audio rendering over heterogeneous devices, in HTML5," in *Proc. Web Audio Conf.*, 2016, pp. 1–6.
[79] S. Landström, J. Bergström, E. Westerberg, and D. Hammarwall, "NB-IoT: A sustainable technology for connecting billions of devices," *Ericsson Technol. Rev.*, vol. 93, no. 3, pp. 1–12, 2016.
[80] V. Lazzarini, D. Keller, C. Kuhn, M. S. Pimenta, and J. Timoney, "Prototyping of ubiquitous music ecosystems," *J. Cases Inf. Technol.*, vol. 17, no. 4, p. 13, 2015.
[81] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 8595–8598.
[82] S. W. Lee, A. D. de Carvalho, Jr., and G. Essl, "Understanding cloud support for the audience participation concert performance of *crowd in c [loud]*," in *Proc. Conf. New Interfaces Musical Expression*, 2016, pp. 176–181.
[83] S. W. Lee, G. Essl, and Z. M. Mao, "Distributing mobile music applications for audience participation using mobile ad-hoc network (MANET)," in *Proc. Conf. New Interfaces Musical Expression*, 2014, pp. 533–536.
[84] G. E. Lewis, "Too many notes: Computers, complexity and culture in *voyager*," *Leonardo Music J.*, vol. 10, no. 10, pp. 33–39, 2000.
[85] F. Lind and A. MacPherson, "Soundtrap: A collaborative music studio with Web audio," in *Proc. Web Audio Conf.*, 2017, pp. 1–2.
[86] "Namm 2013: Panel discussion: Past, present and future of midi," *Future Music Mag.*, 2013, Accessed: Sep. 23, 2017. [Online]. Available: https://www.musicradar.com/news/tech/namm-2013-video-panel-discussion-past-present-and-future-of-midi-570682
[87] J. Malloch, S. Sinclair, and M. M. Wanderley, "A network-based framework for collaborative development and performance of digital musical instruments," in *Computer Music Modeling and Retrieval. Sense of Sounds*. Berlin, Germany: Springer, 2008, pp. 401–425.
[88] J. Malloch, S. Sinclair, and M. M. Wanderley, "Libmapper: (A library for connecting things)," in *Proc. Extended Abstr. Hum. Factors Comput. Syst.*, 2013, pp. 3087–3090.
[89] J. Malloch, S. Sinclair, and M. M. Wanderley, "Distributed tools for interactive design of heterogeneous signal networks," *Multimedia Tools Appl.*, vol. 74, no. 15, pp. 5683–5707, 2015.
[90] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.
[91] O. Mazhelis, E. Luoma, and H. Warma, "Defining an Internet-of-Things ecosystem," in *Internet of Things, Smart Spaces, and Next Generation Networking*. Berlin, Germany: Springer, 2012, pp. 1–14.
[92] D. Mazzanti, V. Zappi, D. G. Caldwell, and A. Brogni, "Augmented stage for participatory performances," in *Proc. Conf. New Interfaces Musical Expression*, 2014, pp. 29–34.
[93] A. McPherson and V. Zappi, "An environment for submillisecond-latency audio and sensor processing on BeagleBone black," in *Proc. Audio Eng. Soc. Conv.*, 2015, pp. 1–7.
[94] C. M. Medaglia and A. Serbanati, "An overview of privacy and security issues in the Internet of Things," in *The Internet of Things*. New York, NY, USA: Springer, 2010, pp. 389–395.
[95] J. B. Meisel and T. S. Sullivan, "The impact of the Internet on the law and economics of the music industry," *Info*, vol. 4, no. 2, pp. 16–22, 2002.
[96] L. Men and N. Bryan-Kinns, "LeMo: Supporting collaborative music making in virtual reality," in *Proc. IEEE VR Workshop Sonic Interact. Virtual Environ.*, 2018.
[97] A. Migicovsky, J. Scheinerman, and G. Essl, "MoveOSC—Smart watches in mobile music performance," in *Proc. Int. Joint Comput. Music Conf. Sound Music Comput. Conf.*, 2014, pp. 692–696.
[98] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, "Internet of Things: Vision, applications and research challenges," *Ad Hoc Netw.*, vol. 10, no. 7, pp. 1497–1516, 2012.
[99] E. R. Miranda and M. M. Wanderley, *New Digital Musical Instruments: Control and Interaction Beyond the Keyboard*, vol. 21. Middleton, WI, USA: A-R Editions, 2006.
[100] J. Miranda *et al.*, "From the Internet of Things to the Internet of people," *IEEE Internet Comput.*, vol. 19, no. 2, pp. 40–47, Mar. 2015.
[101] T. Mitchell, S. Madgwick, S. Rankine, G. S. Hilton, A. Freed, and A. R. Nix, "Making the most of Wi-Fi: Optimisations for robust wireless live music performance," in *Proc. Conf. New Interfaces Musical Expression*, 2014, pp. 251–256.
[102] K. Ng, "Sensing and mapping for interactive performance," *Org. Sound*, vol. 7, no. 2, pp. 191–200, 2002.
[103] R. Oda and R. Fiebrink, "The global metronome: Absolute tempo sync for networked musical performance," in *Proc. Conf. New Interfaces Musical Expression*, 2016, pp. 1–6.

[104] M. Patel and J. Wang, "Applications, challenges, and prospective in emerging body area networking technologies," *IEEE Wireless Commun.*, vol. 17, no. 1, pp. 80–88, Feb. 2010.

[105] D. Poirier-Quinot, B. Matuszewski, N. Schnell, and O. Warusfel, "Nü Soundworks: Using spectators smartphones as a distributed network of speakers and sensors during live performances," in *Proc. Web Audio Conf.*, 2017, pp. 1–7.

[106] I. Poupyrev *et al.*, "Augmented groove: Collaborative jamming in augmented reality," in *Proc. ACM SIGGRAPH Conf. Abstr. Appl.*, 2000, p. 77.

[107] Y. Raimond, S. A. Abdallah, M. B. Sandler, and F. Giasson, "The music ontology," in *Proc. Int. Soc. Music Inf. Retr. Conf.*, 2007, pp. 1–6.

[108] A. R. Netto, L. Castheloge, A. Oliosi, A. Mateus, L. Costalonga, and D. Coura, "Memory tree: Multimedia interactive installation," in *Proc. XV Brazilian Symp. Comput. Music (SBCM)*, Campinas, Brazil, Nov. 2015.

[109] T. Robertson and I. Wagner, "CSCW and the Internet of Things," in *Proc. Eur. Conf. Comput. Supported Cooperat. Work*, 2015, pp. 285–294.

[110] Y. Rogers, H. Sharp, and J. Preece, *Interaction Design: Beyond Human-Computer Interaction*. Hoboken, NJ, USA: Wiley, 2011.

[111] R. Roman, P. Najera, and J. Lopez, "Securing the Internet of Things," *Computer*, vol. 44, no. 9, pp. 51–58, Sep. 2011.

[112] C. Rottondi, M. Buccoli, M. Zanoni, D. Garao, G. Verticale, and A. Sarti, "Feature-based analysis of the effects of packet delay on networked musical interactions," *J. Audio Eng. Soc.*, vol. 63, no. 11, pp. 864–875, 2015.

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

[114] C. Rowland, E. Goodman, M. Charlier, A. Light, and A. Lui, *Designing Connected Products: UX for the Consumer Internet of Things*. Newton, MA, USA: O'Reilly, 2015.

[115] S. Roy, D. Sarkar, S. Hati, and D. De, "Internet of music things: An edge computing paradigm for opportunistic crowdsensing," *J. Supercomput.*, pp. 1–33, Aug. 2018. [Online]. Available: https://link.springer.com/article/10.1007/s11227-018-2511-6#citeas

[116] P. Saari, G. Fazekas, T. Eerola, M. Barthet, O. Lartillot, and M. Sandler, "Genre-adaptive semantic computing and audio-based modelling for music mood annotation," *IEEE Trans. Affect. Comput.*, vol. 7, no. 2, pp. 122–135, Apr. 2016.

[117] A. Sarasúa, "Context-aware gesture recognition in classical music conducting," in *Proc. ACM Int. Conf. Multimedia*, 2013, pp. 1059–1062.

[118] M. Satyanarayanan, "Pervasive computing: Vision and challenges," *IEEE Pers. Commun.*, vol. 8, no. 4, pp. 10–17, Aug. 2001.

[119] M. Satyanarayanan, "The emergence of edge computing," *Computer*, vol. 50, no. 1, pp. 30–39, 2017.

[120] A. Schmeder, A. Freed, and D. Wessel, "Best practices for open sound control," in *Proc. Linux Audio Conf.*, 2010, pp. 1–10.

[121] S. Serafin, C. Erkut, J. Kojs, N. C. Nilsson, and R. Nordahl, "Virtual reality musical instruments: State of the art, design principles, and future directions," *Comput. Music J.*, vol. 40, no. 3, pp. 22–40, 2016.

[122] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

[123] H. Shokri-Ghadikolaei *et al.*, "Millimeter wave cellular networks: A MAC layer perspective," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3437–3458, Oct. 2015.

[124] H. Shokri-Ghadikolaei, C. Fischione, P. Popovski, and M. Zorzi, "Design aspects of short-range millimeter-wave networks: A MAC layer perspective," *IEEE Netw.*, vol. 30, no. 3, pp. 88–96, May 2016.

[125] N. Shrestha, S. Kubler, and K. Främling, "Standardized framework for integrating domain-specific applications into the IoT," in *Proc. IEEE Int. Conf. Future Internet Things Cloud*, Aug. 2014, pp. 124–131.

[126] S. Skach, A. Xambó, L. Turchet, A. Stolfi, R. Stewart, and M. Barthet, "Embodied interactions with E-textiles and the Internet of sounds for performing arts," in *Proc. Int. Conf. Tangible, Embedded, Embodied Interact.*, 2018, pp. 80–87.

[127] M. Slaney, "Semantic-audio retrieval," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, May 2002, pp. 4108–4111.

[128] M. Slater, "Grand challenges in virtual environments," *Frontiers Robot. AI*, vol. 1, p. 3, May 2014.

[129] A. Stolfi, M. Barthet, F. Goródscy, A. Deusany, and F. Iazzetta, "Open band: Audience creative participation using Web audio synthesis," in *Proc. Web Audio Conf.*, 2017, pp. 1–7.

[130] J. Strähle and L. Köhneke, "How digital changed the music industry," in *Fashion & Music*. Singapore: Springer, 2018, pp. 201–221.

[131] A. Subramaniam and M. Barthet, "Mood visualiser: Augmented music visualisation gauging audience arousal," in *Proc. Int. Audio Mostly Conf.*, 2017, pp. 5:1–5:8.

[132] T. Subramanya, L. Goratti, S. N. Khan, E. Kafetzakis, I. Giannoulakis, and R. Riggio, "A practical architecture for mobile edge computing," in *Proc. IEEE Conf. Netw. Function Virtualization Softw. Defined Netw.*, Nov. 2017, pp. 1–4.

[133] A. Tanaka, "Musical performance practice on sensor-based instruments," *Trends Gestural Control Music*, vol. 13, pp. 389–405, 2000. [Online]. Available: https://www.researchgate.net/profile/Atau_Tanaka/publication/239016748_Musical_Performance_Practice_on_Sensor-based_Instruments/links/5707a2e508ae8883a1f7e4ad.pdf

[134] A. Tanaka, "Mobile music making," in *Proc. Conf. New Interfaces Musical Expression*, 2004, pp. 154–156.

[135] F. Thalmann, A. Perez-Carrillo, G. Fazekas, G. A. Wiggins, and M. B. Sandler, "Navigating ontological structures based on feature metadata with the semantic music player," in *Proc. Int. Soc. Music Inf. Retr. Conf.*, 2015, pp. 1–3.

[136] L. Thompson, C. Cannam, and M. B. Sandler, "Piper: Audio feature extraction in browser and mobile applications," in *Proc. Web Audio Conf.*, 2017, pp. 1–3.

[137] L. Turchet and M. Barthet, "An Internet of musical things architecture for performers-audience tactile interactions," in *Proc. Digit. Music Res. Netw. Workshop*, 2017, p. 1.

[138] L. Turchet and M. Barthet, "Envisioning smart musical haptic wearables to enhance performers' creative communication," in *Proc. Int. Symp. Comput. Music Multidiscipl. Res.*, 2017, pp. 538–549.

[139] L. Turchet, M. Benincaso, and C. Fischione, "Examples of use cases with smart instruments," in *Proc. Audio Mostly Conf.*, 2017, pp. 47:1–47:5.

[140] L. Turchet, C. Fischione, and M. Barthet, "Towards the Internet of musical things," in *Proc. Sound Music Comput. Conf.*, 2017, pp. 13–20.

[141] L. Turchet, A. McPherson, and M. Barthet, "Co-design of a smart Cajón," *J. Audio Eng. Soc.*, vol. 66, no. 4, pp. 220–230, 2018.

[142] L. Turchet, A. McPherson, and M. Barthet, "Real-time hit classification in a smart Cajón," *Frontiers ICT*, vol. 5, no. 16, Jul. 2018.

[143] L. Turchet, A. McPherson, and C. Fischione, "Smart instruments: Towards an ecosystem of interoperable devices connecting performers and audiences," in *Proc. Sound Music Comput. Conf.*, 2016, pp. 498–505.

[144] R. van Kranenburg and A. Bassi, "IoT challenges," *Commun. Mobile Comput.*, vol. 1, no. 1, p. 9, 2012.

[145] G. Volpe *et al.*, "Designing multimodal interactive systems using EyesWeb XMI," in *Proc. Conf. New Interfaces Musical Expression*, 2007, pp. 49–56.

[146] J. Wagner, F. Lingenfelser, T. Baur, I. Damian, F. Kistler, and E. André, "The social signal interpretation (SSI) framework: Multimodal signal processing and recognition in real-time," in *Proc. ACM Int. Conf. Multimedia*, 2013, pp. 831–834.

[147] R. H. Weber, "Internet of Things—Need for a new legal environment?" *Comput. Law Secur. Rev.*, vol. 25, no. 6, pp. 522–527, 2009.

[148] R. H. Weber, "Internet of Things: Privacy issues revisited," *Comput. Law Secur. Rev.*, vol. 31, no. 5, pp. 618–627, 2015.

[149] G. Weinberg, "The aesthetics, history and future challenges of interconnected music networks," in *Proc. Int. Comput. Music Conf.*, 2002, pp. 1–8.

[150] G. Weinberg, "Interconnected musical networks: Toward a theoretical framework," *Comput. Music J.*, vol. 29, no. 2, pp. 23–39, Jun. 2005.

[151] M. Weiser, "The computer for the 21st century," *Sci. Amer.*, vol. 265, no. 3, pp. 94–105, 1991.

[152] N. Weitzner, J. Freeman, S. Garrett, and Y. L. Chen, "massMobile—An audience participation framework," in *Proc. Conf. New Interfaces Musical Expression*, 2012, pp. 21–23.

[153] W. Weng, P. Chen, S. He, X. Sun, and H. Peng, "Smart electronic textiles," *Angew. Chem. Int. Ed.*, vol. 55, no. 21, pp. 6140–6169, 2016.

[154] S. Abdallah, E. Benetos, N. Gold, S. Hargreaves, T. Weyde, and D. Wolff, "The digital music lab: A big data infrastructure for digital musicology," *J. Comput. Cult. Herit.*, vol. 10, no. 1, pp. 2:1–2:21, Jan. 2017. [Online]. Available: https://dl.acm.org/citation.cfm?id=2983918

[155] A. Whitmore, A. Agarwal, and L. Da Xu, "The Internet of Things—A survey of topics and trends," *Inf. Syst. Frontiers*, vol. 17, no. 2, pp. 261–274, 2015.

[156] A. Willig, "Recent and emerging topics in wireless industrial communications: A selection," *IEEE Trans. Ind. Informat.*, vol. 4, no. 2, pp. 102–124, May 2008.

[157] T. Wilmering, F. Thalmann, and M. B. Sandler, "Towards a framework for the discovery of collections of live music recordings and artefacts on the semantic Web," in *Proc. Web Audio Conf.*, 2017.

[158] M. Wilson, "Six views of embodied cognition," *Psychonomic Bull. Rev.*, vol. 9, no. 4, pp. 625–636, 2002.
[159] W. Woszczyk, J. Cooperstock, J. Roston, and W. Martens, "Shake, rattle, and roll: Getting immersed in multisensory, interactive music via broadband networks," *J. Audio Eng. Soc.*, vol. 53, no. 4, pp. 336–344, 2005.
[160] M. Wright, A. Freed, and A. Momeni, "Opensound control: State of the art 2003," in *Proc. Conf. New Interfaces Musical Expression*, 2003, pp. 153–160.
[161] Y. Wu, L. Zhang, N. Bryan-Kinns, and M. Barthet, "Open symphony: Creative participation for audiences of live music performances," *IEEE Multimedia Mag.*, vol. 24, no. 1, pp. 48–62, Jan. 2017.
[162] S. Yi, C. Li, and Q. Li, "A survey of fog computing: Concepts, applications and issues," in *Proc. Workshop Mobile Big Data*, 2015, pp. 37–42.
[163] L. Zhang, Y. Wu, and M. Barthet, "A Web application for audience participation in live music performance: The open symphony use case," in *Proc. Conf. New Interfaces Musical Expression*, 2016, pp. 170–175.

LUCA TURCHET received the master's degrees *(summa cum laude)* in computer science from the University of Verona, in classical guitar and composition from the Music Conservatory of Verona and in electronic music from the Royal College of Music of Stockholm, and the Ph.D. degree in media technology from Aalborg University Copenhagen. He is currently a Marie-Curie Post-Doctoral Research Fellow with the Center for Digital Music, Queen Mary University of London, and the Co-Founder and the Head of sound and interaction design at MIND Music Labs. His scientific, artistic, and entrepreneurial research has been supported by numerous grants from different funding agencies, including the European Commission, the Italian Minister of Foreign Affairs, and the Danish Research Council. His main research interests include music technology, Internet of Things, human–computer interaction, and multimodal perception.

CARLO FISCHIONE received the Laurea degree *(summa cum laude)* in electronic engineering and the Ph.D. degree in electrical and information engineering from the University of LAquila. He was a Visiting Professor with the Massachusetts Institute of Technology, an Associate Professor with Harvard University, and a Visiting Scholar and a Research Associate with the University of California at Berkeley. Meanwhile, he has also offered his advice as a Consultant to numerous technology companies, such as the Berkeley Wireless Sensor Network Laboratory, Ericsson Research, Synopsys, and the United Technology Research Center. He is the Co-Funder and the Scientific Director of MIND Music Labs. He is currently an Associate Professor with the ACCESS Linnaeus Center, School of Electrical Engineering, KTH Royal Institute of Technology. He has co-authored over 150 publications, including a book and international patents. His research interests include optimization with applications to networks, wireless and sensor networks, networked control systems, security and privacy. He received a number of awards, including the IEEE Communication Society Stephen O. Rice Award for the best IEEE TRANSACTIONS ON COMMUNICATIONS publication of 2015, the Best Paper Award from the IEEE Transactions on Industrial Informatics, the Best Paper Award at the IEEE International Conference on Mobile Ad-hoc and Sensor System 2005 and 2009, the Best Paper Award from the IEEE Sweden VT-COM-IT Chapter, the Best Business Idea Awards from VentureCup East Sweden and from Stockholm Innovation and Growth Life Science in Sweden, and the Junior Research Award from the Swedish Research Council. He has chaired or served as a technical member of program committees of several international conferences and is serving as a referee for technical journals. He is an Editor of the IEEE Transactions on Communications and an Associate Editor of IFAC *Automatica*.

GEORG ESSL (S'93–M'02) received the bachelor's degree in telematics from the Graz University of Technology in 1996, and the Ph.D. degree from Princeton University in 2002 working with P. Cook on the real-time sound synthesis method for solid objects. He was with the MIT Media Lab Europe and Deutsche Telekom Laboratories, Technical University of Berlin. He has been on the Faculty of the University of Michigan and the University of Florida. He is currently a Visiting Research Professor with the College of Letters and Science, University of Wisconsin–Milwaukee. His research interests are computer music, mobile HCI and mobile music making, human–computer interfaces, real-time physical simulation of audio, tactile feedback, and foundations of numerical methods for interactive applications. He is a member of ASA, ACM, AMS, and ICMA, and serves on the Advisory Board of the New Interfaces for Musical Expression Conference.

DAMIÁN KELLER received the D.M.A. degree from Stanford University in 2004 and the M.F.A. degree Simon Fraser University in 1999. He is currently an Associate Professor with the Federal University of Acre, Brazil, where he coordinates the Amazon Center for Music Research (Núcleo Amazônico de Pesquisa Musical). He is a member and the Co-Founder of the Ubiquitous Music Group (g-ubimus). He has co-authored the volumes *Ubiquitous Music* (Springer) and *Musical Creation and Technologies* (ANPPOM Press). His musical output features the application of eco-compositional techniques in theater, film, electroacoustic, and installation artworks. His research interests include ecologically grounded creative practice and ubiquitous music.

MATHIEU BARTHET is currently a Lecturer in digital media and the Technical Director of the Media and Arts Technology Studios, Queen Mary University of London. His research lies at the intersections of music information retrieval, perception, and human–computer interaction. He currently investigates new interfaces for musical expression, listening, and participatory performance and also conducts fundamental research on musical timbre and emotions. He is a Co-Investigator of the EU Audio Commons Project and a Principal Investigator of the EU project—Towards the Internet of Musical Things.