IOP Conference Series: Materials Science and Engineering

Advanced Experience of Music through 5G Technologies

To cite this article: Adriano Baratè *et al* 2018 *IOP Conf. Ser.: Mater. Sci. Eng.* **364** 012021

Florence Heri-Tech – The Future of Heritage Science and Technologies IOP Publishing
IOP Conf. Series: Materials Science and Engineering **364** (2018) 012021 doi:10.1088/1757-899X/364/1/012021

Advanced Experience of Music through 5G Technologies

Adriano Baratè, Goffredo Haus, Luca A. Ludovico

Laboratorio di Informatica Musicale, Dipartimento di Informatica “Giovanni Degli Antoni”,
Università degli Studi di Milano, 20135 Milano, Italy

E-mail: {`adriano.barate,goffredo.haus,luca.ludovico`}`@unimi.it`

Abstract. This paper focuses on new models to enjoy music that will be implementable in a near future thanks to 5G technology. In the last two decades, our research mainly focused on the comprehensive description of music information, where multiple aspects are integrated to provide the user with an advanced multi-layer environment to experience music content. In recent times, the advancements in network technologies allowed a web implementation of this approach through W3C-compliant languages. The last obstacle to the use of personal devices is currently posed by the characteristics of mobile networks, concerning bandwidth, reliability, and the density of devices in an area. Designed to meet the requirements of future technological challenges, such as the Internet of Things and self-driving vehicles, the advent of 5G networks will solve these problems, thus paving the way also for new music-oriented applications. The possibilities described in this work range from bringing archive materials and music cultural heritage to a new life to the implementation of immersive environments for live-show remote experience.

1. Introduction

The domain of music and music-related information embraces heterogeneous forms of description. The goal of providing a comprehensive representation of a single music piece can be achieved by supplying:

- metadata – textual information about the piece, including its title, the list of involved artists, composition/publishing/release dates, etc.;
- logic information – a “virtual” description of score symbols (chords, rests, time indications, articulation signs, etc.) that does not depend on layout details;
- graphical content – notational instances referring to different score editions, each one with its own format and layout;
- audio/video information – recordings of different performances of the same piece, as well as partial takes of a single performance (like in multi-track audio and video);
- music structures – analysis- or musicology-oriented representations of musical entities (such as themes, subjects, etc.) and their relationships within the composition.

In scientific literature, the mentioned aspects are usually referred to as *layers*, as reported, e.g., in [1], [2], [3]. Currently, suitable computer representations of hierarchical multi-layer music descriptions make use of XML-based standard languages; among others, it is worth citing

IEEE 1599, MEI, and MusicXML formats. XML took foothold in the early 2000s thanks to its intrinsic characteristics, which include the possibility of structuring documents in a hierarchical way, extensibility, the availability of free tools for editing, both human and machine readability, and the support offered by most programming languages [4].

In the list above, we have mentioned a number of heterogeneous facets music information is made of. In addition, in the context of multi-layer representations, each layer can host multiple objects, e.g., different interpretations of the same piece, multi-angle and multi-track takes of the same performance, and differently notated and formatted versions of the same score.

The availability of a unique representation environment supporting multi-layer and multi-instance descriptions provides the user with the possibility to select, navigate, and interact with music contents according to their semantic relationships [5]. File formats and computer-based applications that reflect this approach are already in use, but the characteristics of current mobile-network technologies – in terms of bandwidth, availability, and reliability – have been limiting their use to off-line or wired contexts, so far.

In this sense, the imminent advent of 5G networks will introduce a paradigm shift: this new technology is expected to support reliable delivery of multiple and simultaneous high-quality streams. It will be possible to offer an advanced experience of music and music-related content anywhere and anytime on your own mobile device.

This approach will pave the way for a number of innovative – and sometimes unpredictable – applications, which can raise interest under different perspectives. Regarding entertainment, edutainment and cultural dissemination, users will be able to employ their own mobile devices to experience music in a more effective and engaging way. Also content providers should get an economic benefit, since ad hoc materials must be produced and released, and archive contents can be rediscovered and exploited in new ways. As stated in [6], applications dealing with integrated multi-layer content can even introduce new intellectual rights to be economically exploited, such as the *synchronization right*. Finally, hardware and software developers will have to keep pace with the technological progress and foresee user requests. In conclusion, the application of 5G technologies to musical heritage is expected to involve an entire cultural, multimedia and technological ecosystem.

This work is structured as follows: the next section will describe the state of the art, providing an overview of some relevant and already-available initiatives based on the concept of multi-layer representation of music; Section 3 will outline the characteristics of 5G wireless systems; finally, the last sections will give examples of possible 5G applications, and specifically Section 4 will address music archives, whereas Section 5 will discuss the case of live musical performances.

2. Examples of Multi-layer Applications for Music

In multi-layer formats, heterogeneous information is not only collected and organized within a single document, but also put in relationship so as to obtain intra-layer and inter-layer synchronization. Available formats may implement this concept in different ways, but they mainly rely on the concept of music-event identification: the logical description of a piece presents a number of anchors to which the different representations can be hooked.

A trivial but clarifying example is *score following*, namely the process of listening to a music performance and tracking the position in the score. This goal can be easily accomplished in the context of computer-driven performances, e.g., when music symbols are entered in a digital score editor with playback functions, able to produce a synchronized experience on the fly.

Thanks to multi-layer representation formats, such a concept can be considerably extended: first, it can be applied to other information sources (e.g., lyrics and videos); moreover, the audio layer can include any number of already available recordings, provided that synchronization data are present; the score itself can be any digital or digitized notational instance; finally, homogeneous materials (e.g., the current score or audio instance) can be switched and compared

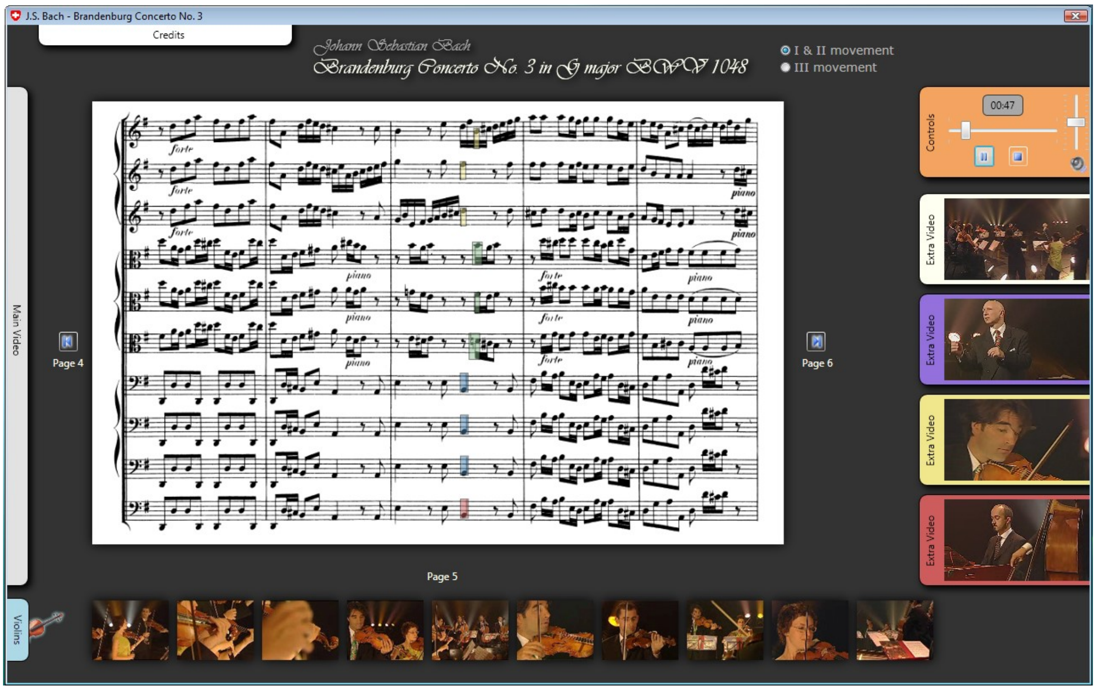


Figure 1. The application to navigate the score of the *Brandenburg Concerto No.3 in G major*. Its interface presents a main area for score following and a scrollable list of videos in the bottom, categorized by instrumental group. Multiple media contents can be played simultaneously in the right panels.

in real time, without affecting the score-following experience.[1]

For the sake of clarity, it is worth mentioning two use cases of multimedia products designed and implemented by the LIM staff and based on a multi-layer music format called IEEE 1599 [7]. The first use case is the prototype realized in 2007 for RSI Radiotelevisione svizzera, in cooperation with the Scuola Universitaria Professionale della Svizzera Italiana (SUPSI). The goal was to provide an advanced score follower for the *Brandenburg Concerto No.3 in G major, BWV 1048* by Johann Sebastian Bach. Its three movements were encoded into IEEE 1599 format with the finest granularity, i.e. allowing note by note synchronization. Music events were mapped onto a single printed score, but synchronized with multiple audio/video streams. The most advanced feature was the availability of about 40 simultaneous video takes, on average 8 per instrumental group; moreover, up to 5 videos could be loaded into ad hoc panels so as to offer a synchronized comparison, as shown in Figure 1. Thanks to this approach, the user could enjoy score following with a very detailed and customizable multi-angle view of the ensemble. Designed as a stand-alone off-line software, the need to keep a high number of video streams synchronized made it impossible – at that time – a web distribution.

Another example of highly demanding application is the multimedia product realized by the LIM in 2011 for the Orchestra sinfonica di Milano Giuseppe Verdi. It had a similar goal, namely advanced score following, but focused on the *Messa da Requiem* by Giuseppe Verdi and presented some additional features. Due to the length of the score and to the high number of orchestral parts to encode, the synchronization was performed at a less refined level of granularity, i.e. measure by measure. This product presented a single full-orchestra score, but 25 simultaneous

---

[1] Please note that the present work mainly focuses on *representation* issues, not on the way synchronized data are obtained, whether manually or automatically, in real time or through off-line processing, etc.

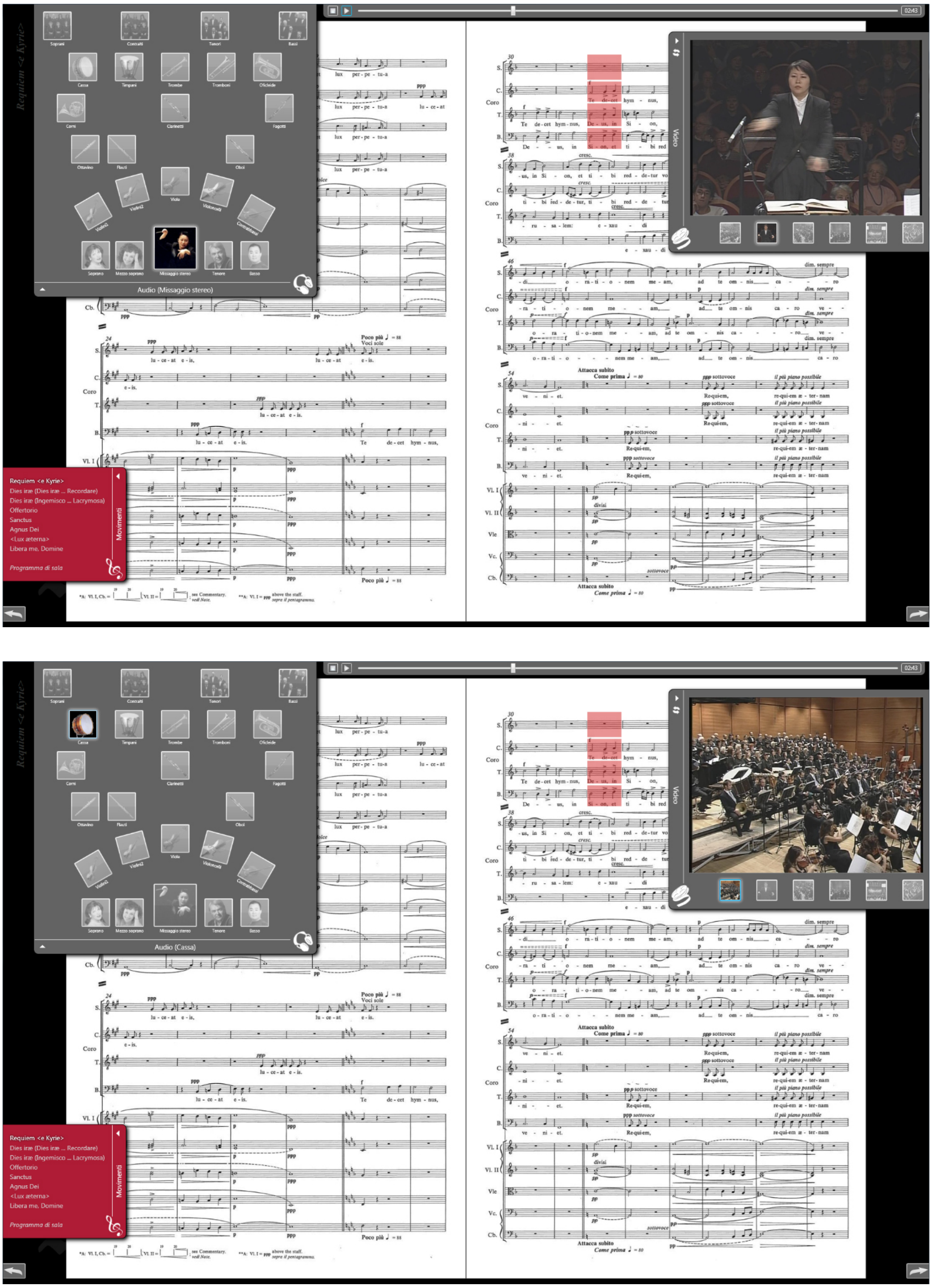


Figure 2. The interface of the *Messa da Requiem* application. The upper figure refers to the selection of the complete audio track and of the camera framing the conductor; in the lower one, the user chooses to listen to the audio from the bass drum microphone and to watch the orchestra from the left side.

audio tracks plus 6 video takes of the event. The most advanced function was the possibility for users not only to customize their visual perspective, like in the RSI prototype, but also to choose one of the multi-track recordings in order to listen to the performance as if they were immersed in the orchestra. The graphical interface of the application is shown in Figure 2.

In the timespan between the release of the two applications, some remarkable innovations were made in the field of network technologies. Defined in 1999, the 802.11b standard operated in the 2.4 GHz band and had a maximum raw data rate of 11 Mbps. The first commercial products supporting this standard appeared on the market in early 2000. An interesting work dating back to that period and addressing the problem of providing application-level quality of service in 3G/4G wireless systems is [8].

The evolution of the standard known as 802.11a was released in 2012, and – operating in the 5 GHz band – it allowed transmission and reception of data at rates of 1.5 to 54 Mbps. Considering congestion and data correction, it yielded realistic net achievable throughput in the mid-20 Mbps. For further details, please refer to [9], a work containing quantitative data about the performance of 3G mobile data offloading through WiFi networks.

Recalling the requirements of our applications, even the WiFi bandwidth would have been sufficient to support high-quality lossy-compressed audio (e.g., 25 audio streams at 320 kbps), but not multiple and simultaneous video content. For this reason, both the applications have been released as stand-alone and off-line products.

In that period, the LIM staff was working on solutions to optimize the delivery of multiple multimedia streams on wired networks, as reported in [10]. A similar fruition on mobile devices seemed far to be achieved. Now, technological advancements in this field are promising, and porting this kind of high-demanding applications onto mobile networks is no longer a chimera.

3. Key Features of 5G Mobile Networks

With respect to the 4G LTE standard, a major step forward is the publication of the "Draft new Report ITU-R M.[IMT-2020.TECH PERF REQ] - Minimum requirements related to technical performance for IMT-2020 radio interface(s)".[2] This document explicits the guidelines for the definition of the 5G standard, the new generation of mobile connectivity that should make its experimental debut in Italy in 2018.

The report establishes the minimum requirements for a mobile antenna to receive 5G certification and, therefore, comply with the specifications of the standard. In some months the draft should be reviewed and confirmed by the companies and research centers that are part of the ITU, the United Nations specialized agency for information and communication technologies. After that step, the draft will become standard and producers will start to release the first 5G-compliant devices.

The draft version of 5G technical specifications establishes at least 20 Gbps downlink and 10 Gbps uplink per single mobile cell. These values are purely theoretical: on one side, telecommunications operators could decide to install antennas with higher peak speeds (e.g., 100 Gbps or more) in order to guarantee better performances for their users; on the other hand, peak velocity refers to the entire cell, and the band – and therefore the connection speed – must be divided among all the users connected at the same time. Realistic transfer rates per single user are expected to be 100 Mbps in download and 50 Mbps in upload. These values are pretty close to the peak speed of some 4G LTE networks, but for 5G this is the average throughput.

Technical specifications fix relevant goals concerning massive connectivity and latency, too. The standard is expected to support at least 1 million connected devices per square kilometer, 1ms to 4ms latency, and the capability to allow up to 500 km/h access (suitable, e.g., for high-speed trains).

---

[2] `https://www.itu.int/md/R15-SG05-C-0040/en`

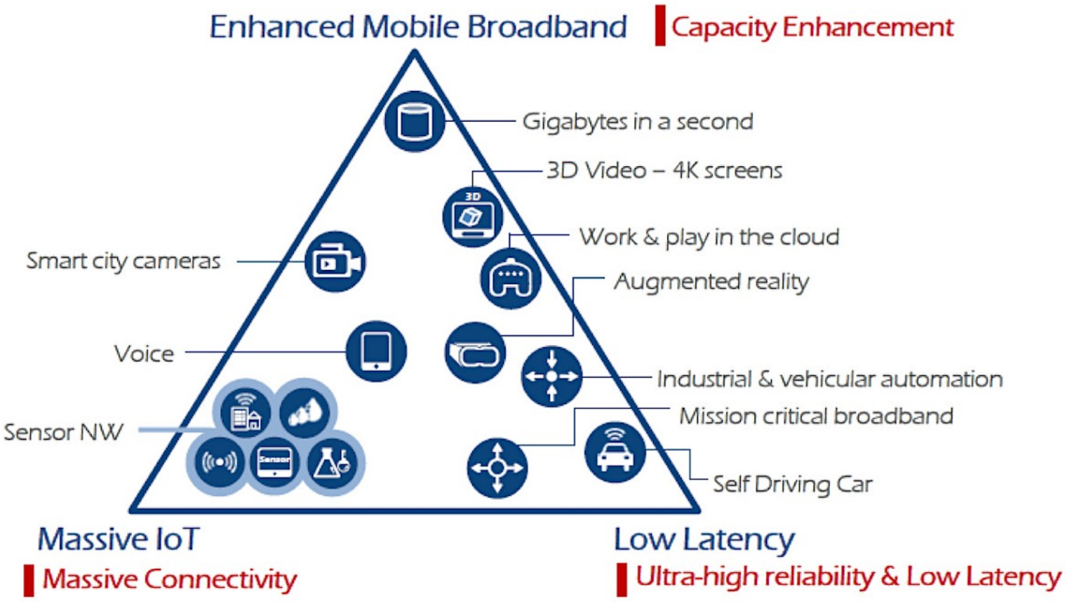


Figure 3. The triangle of 5G applications (source: ETRI graphic, from ITU-R IMT 2020 requirements).

Today, the main mobile services are related to human-centric communication, but the increasing number of devices associated to the Internet of Things (IoT) phenomenon produces forecasts of a total of about 30 billion connected devices by 2020 [11]. This scenario will lead to different communication characteristics, since the type of needs in terms of latency, data rates, density of devices, etc. varies among human-centric communication and IoT devices.

The 5G technology is expected to meet the needs of the next decade [12]. With respect to current network features, it will support:

- 1000 times higher mobile data volume per area;
- 10 to 100 times higher number of connected devices;
- 10 to 100 times higher user data rate;
- 10 times longer battery life for low-power massive machine communication (MMC);
- 5 times reduced end-to-end latency.

A number of scenarios not implementable today will be easily achievable in the future. For instance, Figure 3 outlines a number of possible applications, by arranging them along the axes that link the peculiar features of 5G systems: capacity enhancement, massive connectivity, and ultra-high reliability & low latency.

In [13] these scenarios are described as follows:

- *Amazingly fast* – Users can obtain very high data rates with instantaneous connectivity and low latency. This is crucial for multi-layer applications based on multiple high-quality media streams;
- *Great service in a crowd* – Currently, connectivity is limited when many users share the same area (e.g., stadiums, concert halls, etc.), but in the future also crowded places will permit a satisfactory experience;

Table 1. Video bit rates recommended by YouTube for standard-dynamic-range (SDR) uploads. Values for high-dynamic-range (HDR) videos are similar.

Type	Video Bit Rate Standard Frame Rate (24, 25, 30)	Video Bit Rate High Frame Rate(48, 50, 60)
2160p (4k)	35-45 Mbps	53-68 Mbps
1440p (2k)	16 Mbps	24 Mbps
1080p	8 Mbps	12 Mbps
720p	5 Mbps	7.5 Mbps
480p	2.5 Mbps	4 Mbps
360p	1 Mbps	1.5 Mbps

- *Ubiquitous things communicating* – The mix of IoT and human-centric communications tends to have different needs, and the 5G technology will efficiently handle these new requirements;
- *Best experience follows you* – Even when users are on the move (e.g., traveling by car or commuting), a high quality of service will be guaranteed;
- *Super real-time and reliable connections* – Future wireless systems will support new applications that take full advantage of very high reliability and low latency, thus allowing real-time fruition (e.g., augmented/virtual reality) as well as control (e.g., self-driving vehicles and industrial automation).

Another relevant consequence could be the possibility for some network operators to start deploying so-called *fixed wireless broadband*, i.e. delivering high-speed broadband to areas where there are few or no other broadband suppliers [14].

In order to evaluate the requirements of multi-layer applications, let us consider the bandwidth required by digital media formats commonly in use, focusing on audio and video.

Audio streams for Red Book audio CDs are two-channel signed 16-bit Linear PCM sampled at 44,100 Hz, whose bit rate is 1,411,200 bps, or 176,400 bytes per second. The DVD-Audio format presents a maximum permissible total bit rate of 9.6 Mbps. Concerning compressed audio, files typically present highly variable bit rates, depending on the encoding format, the expected quality, and the characteristics of media content. FLAC files usually have a bit rate that ranges from 220 to 1184 kbps, whereas for MP3 files acceptable values for music applications usually span from 128 to 320 kbps.

Also for video formats there is great variability. At the moment of writing, YouTube adopts an MP4 container with AAC-LC as the audio codec and H.264 as the video codec. Audio playback bit rate is not related to video resolution. Recommended bit rates for audio are: 128 kbps for mono, 384 kbps for stereo, and 512 kbps for 5.1; values for video are shown in Table 1.[3]

Before discussing innovative applications in the field of cultural heritage, let us reconsider the proposals mentioned in Section 2, and specifically a reimplementation that turns those stand-alone softwares into mobile apps interacting with a remote media server. Even leaving latency- and reliability-related issues out, what would the required bit rate be for multiple concurrent media streams?

The *Brandenburg Concerto* application was more demanding as it regards data transfer, due to the high number of selectable video tracks. Please remember that 5G users in metropolitan areas are expected to experience a limit of about 100 Mbps, while 40 full-HD videos encoded in

---

[3] Source: `https://support.google.com/youtube/answer/1722171`

H.264 format with a standard frame rate would require a total bit rate of 320 Mbps. Apparently, despite significant improvements on the bit rate, the 5G technology does not seem ready to support such complex and rich multimedia applications. First, the implications on content quality in order to arrange a given number of media streams may be unacceptable (in the mentioned case, videos could not have a resolution higher than 720p). Secondly, it would be inefficient to send all streams simultaneously to single devices, concerning both network occupation and server throughput.

Luckily, 5G wireless networks offer a solution. Thanks to their very low latency, the client-server request to send a new stream can occur on the fly, with no significant delay perceived in multimedia experience. In other words, it is no more necessary for the client to have a number of media streams simultaneously available and ready to be switched in real time. As a consequence, the band available to the user can be exploited in better ways, e.g. delivering on demand a 4K spherical video that requires 60 Mbps approximatively. Applied to the *Brandenburg Concerto* application, this approach lets the user not only watch the show choosing one of 40 fixed cameras, like in the original version, but also turn around as if he/she was sitting in the middle of the orchestra, in one of the 40 available virtual seatings, and listen to the performance accordingly.

In conclusion, thanks to the impressive features of 5G wireless systems, we can expect the design and release of musical applications not available today, or the reimplementation of already existing ones for personal mobile devices. In the next sections, we will propose two significant use cases focusing on music cultural heritage.

4. Exploitation and Revivification of Music Archives

Our first proposal addresses the exploitation and revivification of archived musical contents.

Until a few years ago, music archives preserved by research centers and international institutions offered very limited access to common users, and sometimes even to professionals. In recent times, thanks to digitization campaigns and web interfaces, many treasures hidden in archives and collections have been unveiled, thus giving new life to intangible cultural heritage.

This kind of cultural dissemination is expected to rise a great interest also in a non-specialized audience, thus becoming a potential source of income for musical institutions, too. Despite the great interest aroused by culture in our days, some critical factors are affecting the phenomenon. First, the user experience offered by specialized applications is often not adequately calibrated on the interests of a general public. In this sense, noteworthy examples are those portals that present lists of uncorrelated digital objects with catalog metadata, maybe useful for a specialized user but totally unappealing for a general audience. Besides, there are technological limitations that prevent a satisfactory fruition on portable devices (screen size and resolution, available memory, etc.), above all when a WiFi connection is not available (limited amount of downloadable data, slow transfer rate, etc.).

The adoption of multi-layer formats is a way to address the former issue, thus capturing the interest of a general public. In fact, by suitably combining, organizing and encoding different information sources, it is possible to achieve multiple results:

- offering a user-tailored experience;
- fostering cultural enrichment through engagement and edutainment applications;
- raising users' interest and extending the audience by launching new multimedia products;
- stimulating the production of ad hoc materials to meet these new needs;
- rediscovering and exploiting the past through novel approaches.

For instance, some historical audio performances have lost their commercial interest, since the availability of new artists (supported and promoted by record labels) and the advancements in recording techniques and formats make new products more appealing for the marketplace. But in a multi-layer environment it is possible, and even desirable, to mix old and new performances into a unique multimedia product.

In the field of education, the revivification of old contents and the production of new ones can be the base for multi-layer applications with a strong valence. For example, the mentioned inter- and intra-layer forms of synchronization typical of a multi-layer environment can encourage learning approaches such as: score-following abilities for performers and conductors, the interpretation of music notations different from the Western traditional one (neumes, tablatures, etc.), the study of vocal repertoire by examples, the comparison of different harmonizations and orchestrations of jazz standards, and so on.

The other mentioned issue, related to technological limitations in the individual fruition of multimedia, will be solved by future advancements in network technologies. Specifically, the features of 5G networks will support any reasonable media payload in all contexts, even in very crowded places. For example, it will be possible to take musicology lessons in a university classroom where students have their own portable devices and anyone interacts autonomously with multiple synchronized media objects. Besides, the characteristics of 5G are fundamental to guarantee the delivering of high-quality multimedia streams with strict time requirements, namely in less than 100ms and aiming at a latency of 5ms, considered acceptable for real-time processing by Digital Audio Workstations.

The adoption of multi-layer formats and the technological support offered by 5G wireless systems could easily bring applications like the ones described in Section 2 into our portable devices. Going one step further, it is not necessary that the content to combine is already available in a single remote repository, rather it could be extracted on the fly from different archives, depending on the user needs and choices.

Let us recall the example of the *Brandenburg Concerto No.3* and re-implement it in an interoperable framework where an agent running on your personal device is able to crawl suitable information sources via the web. Instead of being stored in the limited local disk space, score versions could be provided on-demand by ad hoc repositories (such as IMSLP[4] and Bach Digital[5]), metadata could be retrieved from public digital libraries (e.g., from Europeana[6]), and audio tracks could be extracted from open-source archives (such as the Free Music Archive[7] and the National Jukebox of the U.S. Library of Congress[8]).

With this goal in mind, a relevant problem to solve is the one of standardization and interoperability among different sources. *Interoperability* can be defined as the compatibility of two or more systems such that they can exchange information and data and can use the exchanged information and data without any special manipulation, thus minimizing the loss of information due to technological differences [15]. Thanks to interoperability, users as well as software agents should be able to retrieve data by discovering through a single search what digital objects are available from a variety of collections, rather than having to search each collection individually [16]. From a technical point of view, this problem could be solved through suitable web services exposed by content providers. Unfortunately, another issue arises: collected data, potentially uncorrelated, should be automatically organized into a suitable multi-layer representations. Both goals at present are difficult to achieve.

So far, efforts have focused on the standardization of specific formats for "localized" interoperability. For instance, in 1975 the Cataloguing Commission of the International Association of Music Libraries, Archives and Documentation Centres (IAML)[9] and the

---

[4] `http://imslp.org/`

[5] `https://www.bach-digital.de/`

[6] `https://www.europeana.eu/`

[7] `http://freemusicarchive.org/`

[8] `http://www.loc.gov/jukebox/`

[9] `https://www.iaml.info/`

International Federation of Library Associations and Institutions (IFLA)[10] Committee on Cataloguing drafted the International Standard Bibliographic Description for printed music. The primary goal was to provide the stipulations for compatible descriptive cataloguing worldwide, in order to aid the international exchange of bibliographic records between national bibliographic agencies and throughout the international library and information community [17]. Similarly, in the context of score file interchange, MusicXML was explicitly designed and released to be a universal translator for programs that understand common Western musical notation [18].

Unfortunately, these initiatives – while going in the direction of interoperability – focused only on a subset of characteristics with respect to the potential richness of a multi-layer description. This issue has emerged in recent times in the context of the W3C Music Notation Community Group.[11] Their initial task was to maintain and update the MusicXML and SMuFL[12] specifications in order to develop a standard format for notated music. Starting from the problem of notation interchange over the web, the group soon had to face more general issues, such as enriching scores by arbitrary graphics and coupling them with audio resources such as MP4 and MIDI data, which is the core of the MNX draft.

In a multi-layer ecosystem, a further effort is required to support not only an interoperable exchange of homogeneous information (e.g., catalog metadata, or score files, etc.), but also the compatibility among heterogeneous media types. This problem is solved in multi-layer formats such as IEEE 1599 by providing a common data structure that serves as a dictionary containing the list of labels (ids) to identify each event to be described. After marking an event through an id, multi-layer instances have a non-ambiguous way to hook their descriptions to the event itself. Unfortunately, this approach is hardly feasible in a peer context where different information sources are independent and there is no shared authoritative list of ids.

In principle, synchronization could be achieved in real time by a software agent running on the local device, able to automatically parse logical and multimedia content referring to the same music piece, retrieve synchronization anchors, and feed the media player. Some of these aspects are still open research problems which stand at the intersection of artificial intelligence, pattern recognition, signal processing, and musicology. Examples include reliable optical music recognition [19], complex tasks of music information retrieval [20], and automatic polyphonic score following [21].

Nevertheless, this effort is justified by the potential impact on the revivification and dissemination of musical cultural heritage, and specifically the possibility to enjoy a user-tailored experience on your own device by choosing the multimedia content to combine from a huge set of sources.

5. 5G for Live Musical Performances

Relevant examples of intangible cultural heritages include music and theater, that sometimes merge into a single form of art, such as opera, musical theater, and ballet.

In 2012, a research group at LIM proposed an innovative approach to on-line experience of live performances [22]. The idea was to provide remote viewers, who were watching the show in real time via the web, with a browser application to let them select specific materials from a set of available media contents. For example, the user was allowed to choose the current camera (e.g., watching at the whole scene, focusing on a specific character, or even looking at the audience) and open ad hoc panels carrying additional information (e.g., the script in the case of theatrical performances, or the score for musical ones). A web interface was developed by the LIM to enjoy all the performances of "Prospettiva 09" festival[13] according to this paradigm

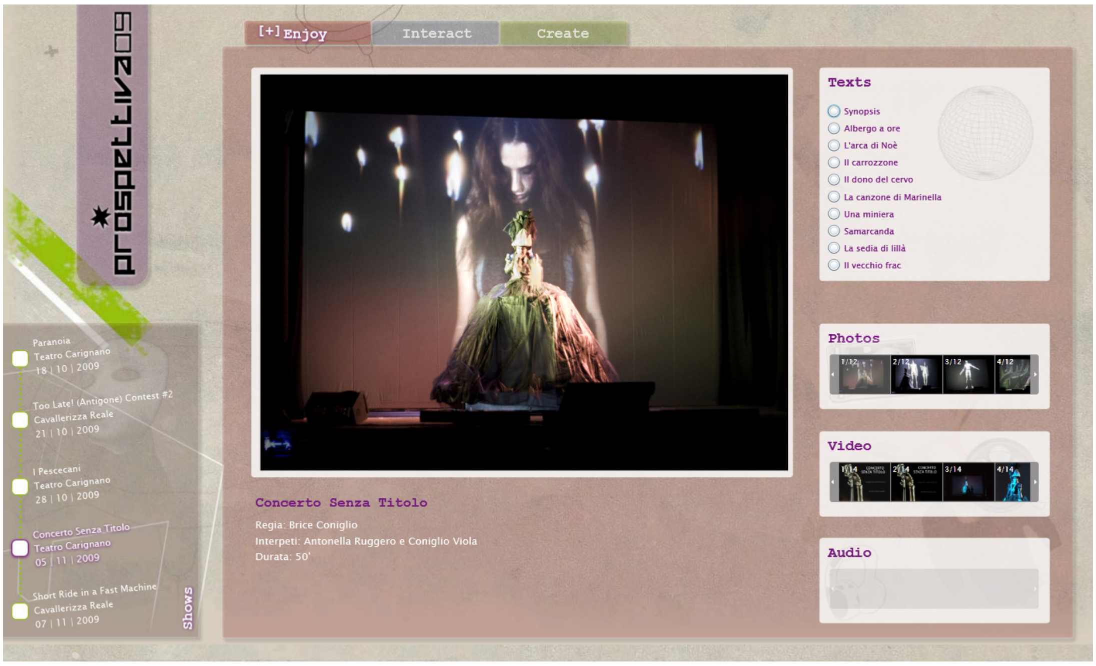
---

[10] `https://www.ifla.org/`

[11] `https://www.w3.org/community/music-notation/`

[12] Standard Music Font Layout, `https://www.smufl.org/`

[13] `http://archivio.teatrostabiletorino.it/collections/occurrence/detail/748/`



Concerto Senza Titolo
Regia: Brice Coniglio
Interpeti: Antonella Ruggero e Coniglio Viola
Durata: 50'

Figure 4. The interface of a web application for the multi-layer fruition of live shows.

(see Figure 4).

The extended interpretation of the concept of *multi-layer representation* implied the availability of multiple audio and video streams, several scripts translated into different languages for theatrical pieces, and different score versions for concerts. The strengths of that proposal were, once again, the concepts of multi-layer fruition, interactivity, and user-tailored experience.

Recalling the examples mentioned in Section 2, there were additional problems to solve, such as multi-cast management for multiple high-quality media streams, the support to real-time user interaction with remote content, and the need for a guaranteed minimum level concerning the quality of service. Due to technological limitations, also this application was intended for high-speed wired connections.

The characteristics of 5G wireless systems illustrated in Section 3 will solve network-related problems, thus supporting the fruition paradigm already explored in a web environment directly on individual devices. But turning a browser app originally conceived for traditional PCs into a mobile app offers further advantages, e.g. the possibility to remotely enjoy music in real time within an immersive virtual-reality (VR) environment. In this way, a plurality of participants may attend an entertainment and/or educational event simultaneously, both in the same physical location and remotely. The performance, delivered in form of spherical video, is intended to be viewed by participants through VR devices such as a head mounted display.

In the case of live musical performances, probably the audience will not be interested in changing the current media stream; nevertheless, additional content could greatly improve user experience. For example, even traditional opera settings are now commonly equipped with devices to follow libretto lyrics (big screens, individual displays, etc.). Through our proposal, it is possible to let the user choose among many translations, select one out of many score versions (all automatically synchronized with current audio), retrieve information about performers, book a new concert date, etc.

A satisfactory user experience implies the delivery of at least one high-quality 360-degree video and multi-channel audio, with a bit rate of about 60 Gbps. Besides, the connection must be reliable and content streaming must be continuous. As mentioned in Section 3, all these requirements will be met by 5G technologies.

Finally, let us mention the possibility to allow remote and geographically-distributed user interaction with the ongoing performance. For example, there are artworks where extemporaneous actions by the audience are encouraged so as to influence the evolution of the plot. It is the case of so-called *immersive theater*, that aims to create participative experiences where audience members give up their "observer" status to become co-actors and co-creators of the narrative and of the storytelling process [23]. The adoption of a multi-layer approach, coupled with the potential of 5G technologies, can foster not only an "augmented" passive experience, but also an interactive participation.

References

[1] Steyn J 2002 Framework for a music markup language *Proceeding of the First International IEEE Conference on Musical Application using* vol 1060 pp 22–29

[2] Haus G and Longari M 2005 A multi-layered, time-based music description approach based on XML *Computer Music Journal* vol 29 (MIT Press) pp 70–85

[3] Ludovico L A 2009 IEEE 1599: A multi-layer approach to music description *Journal of Multimedia* vol 4 (Academy Publisher) pp 9–14 ISSN 1796-2048

[4] Castan G, Good M and Roland P 2001 Extensible markup language (XML) for music applications: An introduction *The Virtual Score* vol 12 pp 95–102

[5] Baratè A and Ludovico L A 2016 Local and global semantic networks for the representation of music information *Journal of e-Learning and Knowledge Society* vol 12 (Italian e-Learning Association (SIe-L)) pp 109–123 ISSN 1826-6223 (print), 1971-8829 (electronic)

[6] Baratè A, Haus G, Ludovico L A and Perlasca P 2016 Managing intellectual property in a music fruition environment - the IEEE 1599 case study *IEEE MultiMedia* vol 23 (IEEE Computer Society) pp 84–94 ISSN 1070-986X

[7] Baggi D L and Haus G M 2013 *Music navigation with symbols and layers: Toward content browsing with IEEE 1599 XML encoding* (John Wiley & Sons)

[8] Fitzek F, Kopsel A, Wolisz A, Krishnam M and Reisslein M 2002 Providing application-level QoS in 3G/4G wireless systems: A comprehensive framework based on multirate CDMA *IEEE Wireless Communications* vol 9 (IEEE) pp 42–47

[9] Lee K, Lee J, Yi Y, Rhee I and Chong S 2013 Mobile data offloading: How much can WiFi deliver? *IEEE/ACM Trans. Netw.* vol 21 (Piscataway, NJ, USA: IEEE Press) pp 536–550 ISSN 1063-6692 URL `http://dx.doi.org/10.1109/TNET.2012.2218122`

[10] Baldan S, Ludovico L A and Mauro D A 2011 Managing multiple media streams in HTML5: the IEEE 1599-2008 case study *Proceedings of the International Conference on Signal Processing and Multimedia Applications (SIGMAP 2011)* ed Linares-Barranco A and Tsihrintzis G A (Seville: SciTePress) pp 193–199 ISBN 978-989-8425-72-0

[11] Ericsson 2015 *Ericsson Mobility Report - On the pulse of the networked society* white Paper URL `https://www.ericsson.com/assets/local/news/2016/03/ericsson-mobility-report-nov-2015.pdf`

[12] Osseiran A, Braun V, Hidekazu T, Marsch P, Schotten H, Tullberg H, Uusitalo M A and Schellman M 2013 The foundation of the mobile and wireless communications system for 2020 and beyond: Challenges, enablers and technology solutions *Vehicular Technology Conference (VTC Spring), 2013 IEEE 77th* (IEEE) pp 1–5

[13] Fallgren M, Timus B *et al.* 2013 Scenarios, requirements and kpis for 5g mobile and wireless system *METIS deliverable D1.1* URL `https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.1_v1.pdf`

[14] Santo B Struggling towards 5G *EDN Network*

[15] Taylor A G and Joudrey D N 2017 *The organization of information* (ABC-CLIO)

[16] Tennant R 2001 Different paths to interoperability *Library Journal* vol 126 (Reed Business Information) pp 118–119

[17] Project Group on ISBD(PM) of the International Association of Music Libraries A and (IAML) D C 1991 *ISBD(PM): International Standard Bibliographic Description for Printed Music* 2nd ed (*UBCIM Publications New Series* vol l) (K. G. Saur Verlag)

[18] Good M and Actor G 2003 Using MusicXML for file interchange *Web Delivering of Music, 2003. 2003 WEDELMUSIC. Proceedings. Third International Conference on* (IEEE) p 153

[19] Rebelo A, Fujinaga I, Paszkiewicz F, Marcal A R, Guedes C and Cardoso J S 2012 Optical music recognition: state-of-the-art and open issues *International Journal of Multimedia Information Retrieval* vol 1 (Springer) pp 173–190

[20] Hu X and Kando N 2017 Task complexity and difficulty in music information retrieval *Journal of the Association for Information Science and Technology* vol 68 (Wiley Online Library) pp 1711–1723

[21] Dorfer M, Arzt A and Widmer G 2016 Towards score following in sheet music images *Proceedings of the 17th International Society for Music Information Retrieval Conference (2016)*

[22] Baratè A, Haus G, Ludovico L A and Mauro D A 2012 IEEE 1599 for live musical and theatrical performances *Journal of Multimedia* vol 7 (Academy Publisher) pp 170–178 ISSN 1796-2048

[23] White G 2012 On immersive theatre *Theatre Research International* vol 37 (Cambridge University Press) pp 221–235