Journal of Music, Technology & Education
Volume 12 Number 1
 doi: 10.1386/jmte.12.1.95_1

ADRIANO BARATÈ, GOFFREDO HAUS, LUCA ANDREA LUDOVICO AND GIORGIO PRESTI
University of Milan

Investigating interpretative models in music through multi-layer representation formats

ABSTRACT

Multi-layer formats are becoming increasingly important in the field of music description. Thanks to their adoption, it is possible to embed into a unique digital document different representations of music contents, multiple in number and potentially heterogeneous in media type. Moreover, these descriptions can be mutually synchronized, thus providing different views of the same information entity with a customizable level of granularity. Standard use cases of multi-layer formats for music address information structuring and support to advanced fruition. The goal of the paper is to demonstrate how suitable multi-layer formats can foster analytical activities in the field of interpretative modelling and expressiveness investigation, discussing both the pedagogical roots and the educational implications of this approach. A use case focusing on the incipit of G. Mahler's Symphony No. 5 *will be presented.*

KEYWORDS

music
informatics
technology
education
multi-layer formats
IEEE 1599

1. Introduction

The technical reproducibility of music, a phenomenon whose origins can be traced back to the second half of the nineteenth century, has dramatically changed musical preservation and listening, causing a paradigm shift in music reception (Arbo 2015). There are a number of reasons why the access to relevant recordings may be of interest for a wide audience, either enthusiasts or experts. First, we can cite the preservation and exploitation of cultural heritage, encouraged by digitization, digital archiving and network technologies (Haus and Ludovico 2006). Moreover, the level of interaction with music language can overtake merely selecting and listening to the desired audio track, achieving multi-modal strategies to retrieve music information (Liem et al. 2011) and cross-modal interaction with content (Damm et al. 2012).

The reproducibility of music has also opened new ways to scientific investigations on performance styles. Performance analysis and interpretative modelling currently represent one of the most relevant fields of computational musicology. The interest in music-performance modelling can have various reasons and goals, ranging from historical research to algorithmic reproduction. In the past, both general and detailed aspects of music interpretation have been deeply investigated in scientific literature: please refer to Shaffer and Todd (1987), Clarke (1988), Palmer (1996), Mazzola and Göller (2002) and Gabrielsson (2003), to cite but a few works. These activities can be supported by the automatic retrieval and analysis of quantitative data from recorded musical performances. As stated in (Widmer and Goebl 2004), the purpose of computational models of expressive music performance is to specify precisely the physical parameters defining a performance (e.g., onset timing, inter-onset intervals, loudness levels, note durations, etc.) and to quantify quasi-systematic relationships among certain properties of the musical score, the performance context and an actual performance of a given piece.

The idea of analysing different performances to unveil interpretative models has been employed in a number of scientific works, such as Friberg and Sundströöm (2002), Goebl et al. (2004) and Repp (1990). One of the major problems for an extensive analysis of multiple audio tracks is to determine matches among occurrences of audio events in the various media. On one side, an automatic approach to synchronization would be desirable, but its result could be roughly approximated or even misleading; on the other side, a manual or supervised approach would be more reliable from a musicological point of view, but time-consuming and potentially less accurate on specific parameters (e.g., the exact timing of note attacks).

The idea proposed in this work is to exploit the potential of multi-layer music representation formats to extract both synchronization information (i.e. *where* music events occur in different music tracks) and logic information (i.e. *what* is the definition of such events from a symbolic point of view). Please note that the approach presented in this paper will be descriptive rather than predictive: we will focus on measuring performance details and describing classes of common patterns with the help of statistical analysis, rather than trying to infer a model whose predictions have to be compared to real performances.

The paper is organized as follows: Section 2 provides an overview about multi-layer formats in education, Section 3 applies these concepts to the case of music, Section 4 addresses the core problem of audio synchronization, Section 5 exemplifies automatic feature extraction from a suitable multi-layer

format, Section 6 presents an explanatory case study and Section 7 presents final remarks concerning implications in music education.

2. Pedagogical roots for a multi-layer educational approach

This work has its pedagogical roots in the theory of interactive multimedia and multimodal learning environments, widely discussed in the literature (Najjar 1996; Cairncross and Mannion 2001; Mayer 2002; Moreno and Mayer 2007; Sankey et al. 2010).

The key idea of multi-layer formats is to provide a comprehensive description of a given information entity by catching its multiple aspects. This approach is commonly in use, e.g., in augmented reality applications, where the view of a physical, real-world environment is enriched by additional information – including sound, video, graphics, haptics, etc. – to offer an extended experience of a situation or a better comprehension of a phenomenon.

Moreover, for each type of content, multi-layer environments can present a multiplicity of information objects, e.g., multiple text descriptions or audio files. In the following, these two levels of abstraction (i.e. the content type and the specific object of that type) will be defined as *layer* and *instance*, respectively.

In conclusion, a multi-layer format aims to describe a single information entity through multiple (heterogeneous) layers, each one potentially carrying multiple (homogeneous) instances. Instances may present a number of relationships towards other instances within the same layer (*intra-layer synchronization*) or instances from other layers (*inter-layer synchronization*). Some clarifying examples from the music domain will be presented in Section 4.

In education, multi-layer environments can be very effective to foster a deeper understanding of learning objects, analysing them from a number of perspectives. In this sense, one of the pedagogical theories of reference is *example-based learning* (Atkinson and Renkl 2007), as the availability of multiple homogeneous instances, mutually connected, can encourage the processes of learning and abstraction. Example-based learning, which in music can be translated as playing and learning by ear, is commonly in use, as Priest (1989), Woody (2012) and Green (2017) remark in their works, to mention but a few. This aspect is sometimes abused: suffice it to cite the 'historicized' variations of the leading part of famous arias originated by great performances of the past. Often, opera singers, instead of studying the original score, prefer to learn their part by listening to old recordings.[1] Even if example-based learning is subject to misuse in music, a multi-layer environment carrying multiple audio instances (including philological ones) and linking them to the right score information could help in preventing bad learning habits.

Other applications of a multi-layer educational approach in music will be discussed in Section 7.

During the design phase of a multimedia learning environment based on a multi-layer approach, it is important to consider the *cognitive load theory* by Sweller (Sweller et al. 2011). Cognitive load refers to the total amount of mental effort required to the working memory by the information presented to learners. Such a theory argues that – when designing training paths – it is essential to start from the cognitive structure of human mind and to pay attention to the conditions of overload to which work memory can be subjected. This theory provides a set of theoretical guidelines useful in didactic design to support the efficiency of the learning process. Some studies – e.g., Paas and

1. An example is the high C sung by Manrico in the cabaletta 'Di quella pira', Act 3, Scene 2 of Giuseppe Verdi's opera *Il trovatore*.

Van Merriënboer (1993) – identified categories of factors that can influence cognitive load, including causal, task-related, environmental and evaluation factors.

In the context of multimedia learning, the adoption of a multi-layer approach could increase the amount of external cognitive load by subtracting energy to the relevant cognitive load, thus affecting student learning outcomes. In these cases, it is important to know the principles and guidelines that domain experts have outlined to design learning experiences that can reduce the external cognitive load, especially in presence of high levels of complexity (Sweller et al. 1998; Van Gog and Paas 2008; Van Merrienboer and Sweller 2005). For further details, please refer to Faiella and Mangione (2012).

3. Multi-layer formats for music description

The goal of providing a comprehensive description of a music work can have multiple meanings and involve different domains. First, the intrinsic content of a music piece can be seen as an organized flow of music or sound events established by the composer. This level of detail is sometimes referred to as the *logic description*. In order to be expressed, the logical description is often transcribed into notated music. Notation is any system that represents scores through the use of written symbols, including modern staff notation, neumes, tablatures, Braille transcriptions, alternative graphical representations, etc. The transcription process makes music symbols instanced in the graphical domain, thus producing different versions of the score. Similarly, as it regards the audio domain, score interpretation by musicians or computer-based systems produces different music performances. Other kinds of information may further enrich the description of a given music work, for instance metadata, lyrics, onstage photos, playbills, etc.

The problem of catching and describing heterogeneity in the digital domain has been traditionally faced through a number of different media formats, each one addressing a specific aspect of music information; for example: binary (e.g., MakeMusic Finale, MuseScore and Avid Sibelius), text-based (e.g., ABC, GUIDO and DARMS) and XML-based (e.g., IEEE 1599, MEI and MusicXML) music-notation formats for logic descriptions; graphic file formats (e.g., JPEG, PNG and TIFF) for graphical score instances; digital audio formats (e.g., AIFF, MP3 and WAV) for sound tracks; computer-driven performance languages (e.g., MIDI, Csound and SASL/SAOL).

A key problem emerging from the adoption of specific formats is the difficulty of relating different descriptions of the same information entity: How to link multiple notated instances – belonging to different score editions – of the same music symbol? How to relate different performances – belonging to different recordings – of the same excerpt? And how to synchronize the cursor advancement over a score to the timed playback of a given audio track, thus implementing score following, allowing to switch both score versions and audio performances on the fly?

An emerging approach is to provide a comprehensive, integrated and synchronized description of music through a single format. A multi-layer structure is suitable to treat complex and rich information by keeping contents properly organized within a unique framework. Providing a multi-layer description implies describing an entity from different perspectives, thus unveiling its heterogeneous facets, and music information is made of heterogeneous facets whose degree of abstraction may range from purely logical descriptions to physical signals.

As reported in scientific literature (Lindsay and Kriechbaum 1999; Steyn 2002; Haus and Longari 2005), different aspects of music can be fully covered by the following layers: *general*, *logic*, *structural*, *notational*, *performance* and *audio*. First, layers can be seen as containers, thus underlining their capability to support and logically organize multiple instances in each layer. In this way, it is possible to associate a given music piece to many logic descriptions (e.g., the original score and multiple revised versions), many printed scores (each one decomposable into a variable number of graphical files) and many audio/video tracks. It is worth pointing out that not all layers have to exist for a given music piece. For instance, jazz music is often based on extemporaneous improvisation and it does not present a standard score: in this case, the audio layer could contain multiple instances, but the logic layer would be missing. Needless to say, the presence of many layers and the availability of many instances within single layers provide a richer description of the music piece, thus allowing a comprehensive experience of music in all its aspects and opening the way for advanced applications oriented to fruition, analysis, education, etc.

Examples of in-use formats for music description that adopt a multi-layer approach are: Music Encoding Initiative (Roland 2002), MusicXML (Good and Actor 2003), MPEG-SMR (Bellini et al. 2005) and IEEE 1599 (Baggi and Haus 2013).

4. Synchronization

In the previous section, we have mentioned the role of containers played by layers in some music formats. This approach may recall the storage features of directories or compressed archives, where a single entity is used to embed a number of correlated documents. Nevertheless, multi-layer formats typically present more advanced peculiarities, for example the possibility to express synchronization among contents.

There are different ways to achieve this result. For example, MIDI explicitly provides syntactic elements to control synchronization in a MIDI chain. Unfortunately, MIDI is not suitable for our purposes: first, this language does not embed or link external audio files, since it was mainly conceived to drive sound-generation modules (i.e. synthesizers) to achieve a computer-based performance; and the adoption of MIDI-based protocols, such as *MIDI Machine Control*, to control MIDI-capable media players would be only a tricky and partial solution. Besides, MIDI does not contain a sufficiently accurate description of music notation, introducing unacceptable simplifications for the purposes of computational musicology (such as the definition of note pitches and durations).

About ten years ago, the MPEG format embedded a so-called Symbolic Music Representation (SMR), namely a logical structure based on: (1) symbolic elements representing audio-visual events, (2) the relationship between those events, and (3) aspects related to how those events can be rendered (visually as music notation or audibly) and synchronized with other media types (Nesi et al. 2006). The official documentation was published under the title 'ISO/IEC FCD 14496-23:200x – Symbolic Music Representation'. For historical and technological reasons, multimedia content is the core of MPEG-SMR vision. In fact, before the integration of music notation modelling, MPEG-4 technology already covered a huge media domain including synthetic and natural hybrid coding (SNHC) audio, semi-symbolic audio and computer-driven

performance languages (like MIDI) and structured descriptions of audio through a normative algorithmic language associated with a score language (MPEG-4 Structured Audio [SA]), as described in (Scheirer 1998). All these contributions can be rendered and synchronized with other forms of media: images, video, graphic animations, etc.

A different approach is the one adopted by the IEEE 1599 standard. Since this format has been explicitly conceived for music description, its core is the logic representation of scores in terms of music symbols. To achieve the synchronization goal, the encoding contains a common data structure called the *spine* that lists and uniquely identifies all the information entities – i.e. *music events* – to be described in other layers. All multimedia descriptions of these entities are demanded to the corresponding layers, which embed one or many media instances (i.e. graphical or audio files), as explained in D'Aguanno and Vercellesi (2007).

For the sake of clarity, let us consider a generic music event e_k:

- e_k can be defined in terms of music notation within the logic layer (e_k could be a C4 quarter note at the beginning of the first measure of the flute part);
- e_k's graphical renditions can be retrieved from the notational layer (e_k could be in a given position of the first page of the manuscript, in a different position on a solo-flute transcription and in another page of a given printed orchestral score);
- e_k's acoustic renditions, produced by different flutists during their musical performances and each one with its own timing, are listed in the audio layer.

IEEE 1599 explicitly supports *intra-layer* and *inter-layer synchronization* (see Section 2), to highlight, respectively, the links among homogeneous and heterogeneous descriptions of the same music event.

Basic interpretative modelling would only require a format able to synchronize multiple audio instances. For example, if we have a number of performances of the same music piece and the modelling goal is to track delta times between couples of contiguous music events comparing them to other performances, synchronization among multiple tracks is sufficient. Please note that the way synchronization anchors are obtained – either automatically, manually, through supervised algorithmic approaches, etc. – is not relevant in this context; in any case, the identification of event timings must be precise and trustworthy.

Nevertheless, the presence of other layers, each one contributing with additional information, can be exploited to implement more detailed kinds of analysis. For instance, a logic description of the piece allows to map sound events onto score symbols. If the modelling goal is, say, to analyse beats-per-minute (BPM) variations over multiple performances, time distances obtainable from the audio layer are not sufficient; rather, they must be related to rhythmical values of music events, an information typically belonging to the logic domain. These considerations can be extended to the presence of additional layers.

This is the core subject of the paper: starting from rich organized data sets – which may imply a higher number of mutually synchronized audio tracks and additional information available in other layers – would improve interpretative modelling and enable more advanced features, including innovative interfaces for the experience and comparison of interpretative models. A multi-layer

approach is fundamental both to build models themselves (e.g., unveiling the conducting style by Arturo Toscanini) and to perform re-synthesis attempts (e.g., applying that conducting style to a computer-driven performance).

Similar initiatives have been already conducted in the field of computational musicology. It is worth citing, for example, the research activities by Repp (1990, 1992), focusing on Beethoven's and Schumann's repertoire, and the *Mazurkas project* of the AHRC Research Centre for the History and Analysis of Recorded Music (CHARM), aiming to investigate the style, performance and meaning in Chopin's Mazurkas (Cook 2007). With respect to such initiatives, the novel aspect of our proposal is the adoption of a multi-layer format, specifically IEEE 1599, to join all needed information and to intrinsically facilitate the analysis and recognition of diversity and commonality in music performance.

5. Automatic feature extraction

As discussed in Section 4, the availability of music data in a reliable synchronized form can foster the automatic extraction of performance-related characteristics.

A non-exhaustive list of music and audio characteristics that can be analysed on each recording to produce an interpretative model may include:

- metronome-related information, i.e. the average value measured on the whole piece or a specific section, or a function that draws the changes occurring in the BPM value;
- instrument tuning and intonation accuracy;
- loudness and volume envelopes;
- spatial properties, such as rough environment reverberation statistics, soloist/orchestral section seating and other acoustic features commonly used in Music Information Retrieval to describe audio content.

So far, we have adopted a multi-layer format basically as a container that embeds a number of performances, an approach that achieves non-trivial results such as note-by-note synchronization. In addition, a multi-layer format can provide further information that may be useful for musicological considerations, e.g., concerning piece and performance metadata (described in the logic and audio layer, respectively). In this way, it is possible to cluster the performances of a given music piece by a number of features, such as:

- By conductor or soloist – To what extent conductors and/or soloists influence the performance and give their imprint to the result? For example, is there a common, automatically detectable root in the four cycles of Beethoven's symphonies that Herbert von Karajan conducted and recorded during his life?
- By orchestra – Does an ensemble lend well recognizable features to a performance, regardless of the conductor? For instance, when Berliner Philharmoniker play in their concert venue, does the orchestra present a peculiar sound (orchestra seating, hall reverberation, etc.)?
- By geographical area, time period, culture and education – Do these factors influence performances so as to make them clearly distinguishable? Are oriental piano players technically precise but unemotional in their performances? Does the Italian *bel canto* school emerge from the analysis of

opera recordings? Are the current executions of Baroque repertoire more philological than in the past?
- By signal properties – Has the environment (say a concert hall, an opera house, a rehearsal room, etc.) a similar impact on pieces or performances with different characteristics? This question may also recall the *acoustic ecology* studies started in the late 1960s with R. Murray Schafer and his team at Simon Fraser University in the context of the World Soundscape Project (Wrightson 2000).

Moreover, MIR algorithms can take advantage of the proposed framework thanks to prior information about what is under investigation. For example, knowing which fundamental frequencies are playing – an information that can be extracted from the logic layer – can wipe away the uncertainty introduced by a blind pitch tracking when estimating features based on harmonicity (Peeters et al. 2011).

It is worth underlining that we are not directly analysing players' performances, but rather the final results of their digitization. From this point of view, some parameters could have been significantly altered, being affected by the upstream activities of recording, mixing/editing and digital mastering. In some cases, this is evident. For example, sound level is potentially influenced by the whole audio signal flow, which includes not only sound sources, but also microphones, preamplifiers, equalizers, compressors and finally analog-to-digital converters. At each step, sound parameters are intrinsically or even deliberately manipulated for artistic or technical purposes. It is worth mentioning the practice called *loudness war* (or *loudness race*), namely the trend of increasing audio levels thus compressing the dynamic range in recorded music during the audio mastering phase (Vickers 2010). Due to the potential occurrence of these operations during audio acquisition, mixing and mastering, it makes no sense to evaluate the peak amplitude of different recordings to compare performances of a *fortissimo* from a full orchestra.

In other cases, sensing the subtle difference between live performance and recording may be harder. For example, a non-expert might consider the instruments tuning or the metronomic values as independent of the recording and playback conditions, but it is sufficient a small change, say, in the steady speed of a turntable or in the DAC clock of a device to alter them.

Since our proposal is based on the availability of materials already in digital form, we will assume that the digitization process has been carried out in the best way. Moreover, in the case study reported below, we will try to limit the effects introduced by audio recording chains.

6. Case study

To demonstrate the potential of multi-layer formats to investigate interpretative models, we encoded in IEEE 1599 format 50 recordings of the 1st movement of G. Mahler's *Symphony No. 5*. The complete list of audio tracks is listed in Table 1, together with the corresponding short labels used in the following figures. For the sake of brevity, our investigation is limited to the trumpet solo contained in the first fourteen measures, as shown in Figure 1.

We focused on two categories of parameters, one related to tempo (a music-related feature) and the other to amplitude perception (an audio-related feature). Needless to say, the same approach could be extended to a complete music score or a set of pieces and could take into account other characteristics.

ID	Conductor, orchestra, recording date (release date, record label)
Abb1973	Abbado, Chicago Symhpony Orchestra, 1973 (1989, DG)
Abb2004	Abbado, Lucerne Festival Orchestra, Aug. 2004 (2005, EuroArts)
Ash2010	Ashkenazy, Sydney Symphony, May 2010 (2010, Sidney Symphony)
Bar1969	Barbirolli, New Philharmonia Orchestra, Jul. 1969 (1988, EMI)
Bar1997	Barenboim, Chicago Symphony Orchestra, Jun. 1997
Ber1972	Bernstein, Wiener Philharmoniker, Apr. 1972 (2005, DG)
Ber1975	Bernstein, Wiener Philharmoniker, 1975 (2010, DG)
Ber1990	Bertini, Kolner Rundfunk-Sinfonieorchester, Jan.–Feb. 1990 (2005, EMI)
Bou1996	Boulez, Wiener Philharmoniker, 1996 (1997, DG)
Bri1998	Briggs, Apr. 1998
Cha1997	Chailly, Royal Concertgebouw Orchestra, Oct. 1997 (1998, Decca)
Dar2010	Darlington, Duisburger Philharmoniker, Sept. 2010 (2011, Acousence)
Dud2006	Dudamel, Simon Bolivar Youth Orchestra of Venezuela, Feb. 2006 (2007, DG)
Esc2009	Eschenbach, Orchestre de Paris, Mar. 2009
Far1980	Farberman, London Symphony Orchestra, 1980 (1999, Vox)
Fis2013	Fischer, Budapest Festival Orchestra, 2013
Gat1997	Gatti, Royal Philharmonic Orchestra, Nov. 1997 (1998, BMG)
Ger2001	Gergiev, London Symphony Orchestra, 2001 (2001, LSO)
Hai1968	Haitink, Concertgebouw Orchestra, 1968 (1994, Decca)
Hai1970	Haitink, Royal Concertgebouw Orchestra, Dec. 1970 (1994, Philips)
Hon2011	Honeck, Pittsburgh Symphony Orchestra, May 2011 (2011, Octavia)
Hor1969	Horenstein, Gothenburg Symphony Orchestra, 1969
Kar1973	von Karajan, Berliner Philharmoniker, Feb. 1973 (1973, Polydor)
Kem1948	Kempe, Rundfunk-Sinfonieorchester Leipzig, Nov. 2948 (2011, Archipel)
Kon1974	Kondrashin, USSR Symphony Orchestra, 1974
Kre2010	Kreizberg, Orchestre Philharmonique de Monte-Carlo, Sept. 2010 (2012, OPMC)
Kub1951	Kubelík, Concertgebouw Orchestra of Amsterdam, Jun. 1951 (2001, Tahra)
Kub1971	Kubelík, Symphonieorchester des Bayerischen Rundfunks, Jan. 1971
Kub1981	Kubelík, Symphonieorchester des Bayerischen Rundfunks, Jun. 1981 (1999, Audite)
Lei1963	Leinsdorf, Boston Symphony, 1963
Lev1977	Levine, Philadelphia Orchestra, 1977 (1978, Sony)
Meh1976	Mehta, Los Angeles Philharmonic Orchestra, 1976
Meh1989	Mehta, New York Philharmonic, Sept.–Oct. 1989 (1990–97, Teldec)
Mit1960	Mitropoulos, Philharmonic-Symphony Orchestra, Jan. 1960 (2011)
Mor1993	Morris, Symphonica of London, 1993 (1993, IMP)
Nan1995	Nanut, Rundfunk Symphony Orchestra Ljubljana, 1995 (1995, Forum)
Neu2004	Neumann, Czech Philharmonic Orchestra, 2004 (2004, EP)

(Continued)

ID	Conductor, orchestra, recording date (release date, record label)
Neu1967	Neumann, Gewandhausorchester Leipzig, 1967 (2006, Edel)
Nor2006	Norrington, Radio-Sinfonieorchester Stuttgart des SWR, Jan. 2006 (2006, Hanssler)
Oza1990	Ozawa, Boston Symphony Orchestra, Sept. 1990 (2002, Decca)
Rat2002	Rattle, Berliner Philharmoniker, Sept. 2002 (2007, EMI)
Roh1973	Roždestvenskij, Moscow Radio Symphony Orchestra, Dec. 1973
Sch1952	Scherchen, Orchester der Wiener Staatsoper, 1952 (2002, DG)
Sin1985	Sinopoli, Philharmonia Orchestra, Jan. 1985
Sol1970	Solti, Chicago Symphony Orchestra, Mar. 1970 (1996, Decca)
Sui2003	Suitner, Staatskapelle Berlin, 2003 (2003, Edel)
Tem2003	Temirkanov, Saint Petersburg Philharmonic Orchestra, Sept. 2003 (2005, Water Lily Acoustics)
Til2005	Tilson Thomas, San Francisco Symphony, Sept.–Otc. 2005 (2006, SFS)
Waa1992	de Waart, Radio Filharmonisch Orkest Holland, 1992
Wal1947	Walter, New York Philharmonic, Feb. 1947 (2012, Sony)

Table 1: List of analysed recordings.

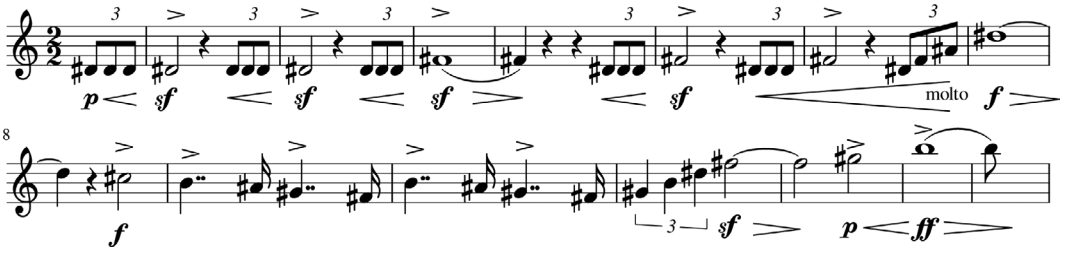

Figure 1: G. Mahler, Symphony No. 5, *1st movement, trumpet-solo part, measures 1–14.*

During the synchronization process, we have determined fifteen anchors particularly easy to discover thanks to the characteristics of waveforms (e.g., note attacks after rests). In the following figures – when needed – we will show the score excerpt vertically aligned below diagrams and indicate anchor points through small circles. First, we have analysed the average BPM for each performance, discovering a great variability in soloists' interpretations. This result is not surprising, since the first movement (*Trauermarsch*, i.e. Funeral march) presents a mood marking with a rough tempo connotation: *In gemessenem Schritt. Streng. Wie ein Kondukt*, which means 'at a measured pace, strict, like a funeral procession'. In this case, since neither a commonly recognized tempo marking nor an explicit metronome are suggested, the BPM values noticeably change from a performance to another. The automatic processing of synchronization timings in the audio layer pointed out a range between 72 and 156 BPM at beat 4, and between 50 and 138 BPM at the end of the solo, as shown in Figure 2.

Considering only the beginning and the end of the solo, these values highlight an average slowdown in the execution speed, probably due to the entrance of the orchestra at beat 47. But the analysis of BPM graphs points out

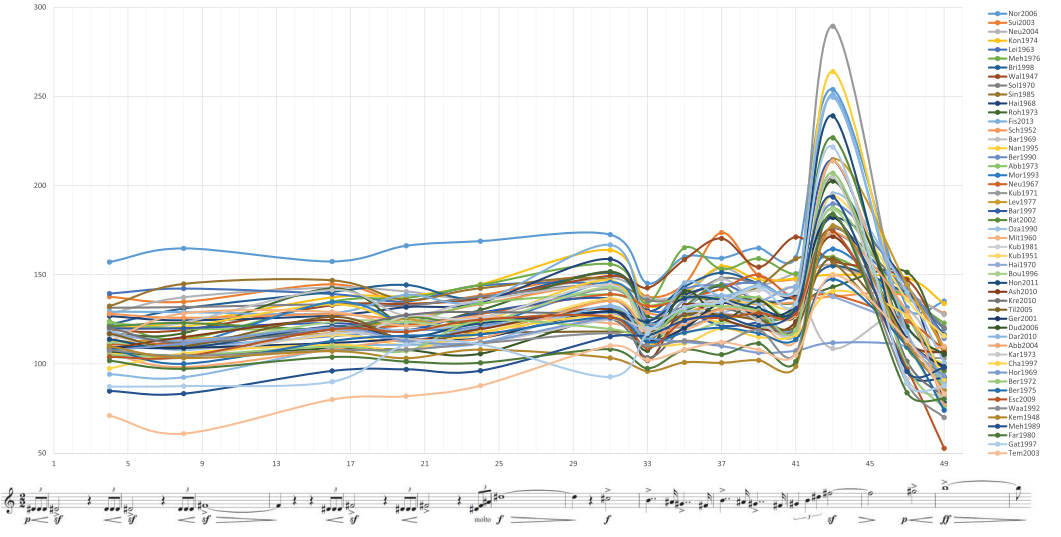

Figure 2: BPM as a function of music beats. Coloured lines represent BPM variations track by track, and circles are synchronization anchors referring to the score below.

another clearly recognizable behaviour: the triplet of quarter notes starting at beat 41 is commonly performed with a gradual acceleration, and it is rare to find a performance perfectly timed (e.g., Hor1969) or even decelerating (e.g., Bar1969). This peak in BPM, reached during the ascension of the melodic line, is usually followed by a sharp slowdown related to the full-orchestra cadence. From a musicological point of view, the two chords of the cadence – occurring at beat 47 and 49, respectively – have a strongly affirmative meaning, and usually communicate to the audience a sense of solemnity that is further underlined by a *rallentando*.

Many other analyses can be automatically performed on tempo-related aspects. For instance, the box plot in the left part of Figure 3 shows the BPM ranges covered measure by measure by all tracks, subdivided into quartiles. The diagram illustrates the measures where tempos converge (short segments) or diverge (long segments). Among the 50 recordings analysed here, which represent a comprehensive testbed, measure 12 once again presents the most variable behaviour. The box plot in the right part shows BPM variations track by track, thus highlighting which soloists were rhythmically regular (short segments) or free (long segments). For instance, in the 1999 concert conducted by Rafael Kubelík and performed by the Symphonieorchester des Bayerischen Rundfunks we notice a spike apparently conflicting with a low average BPM, which leads us to expect a rapid acceleration somewhere in the performance.

Now let us consider *loudness*, namely that characteristic of a sound that is a correlate of physical strength (i.e. amplitude). In this case, a trivial analysis could be the comparison among RMS[2] curves for each track. As mentioned in Section 5, such an analysis can provide information about audio signal flows on different recordings, and not on the original performances themselves. To limit the effects introduced by the audio chain, we first segmented each track incipit into an equal number M of segments, with $M = 14$. Then, for each m-th

2. RMS stands for Root Mean Square.

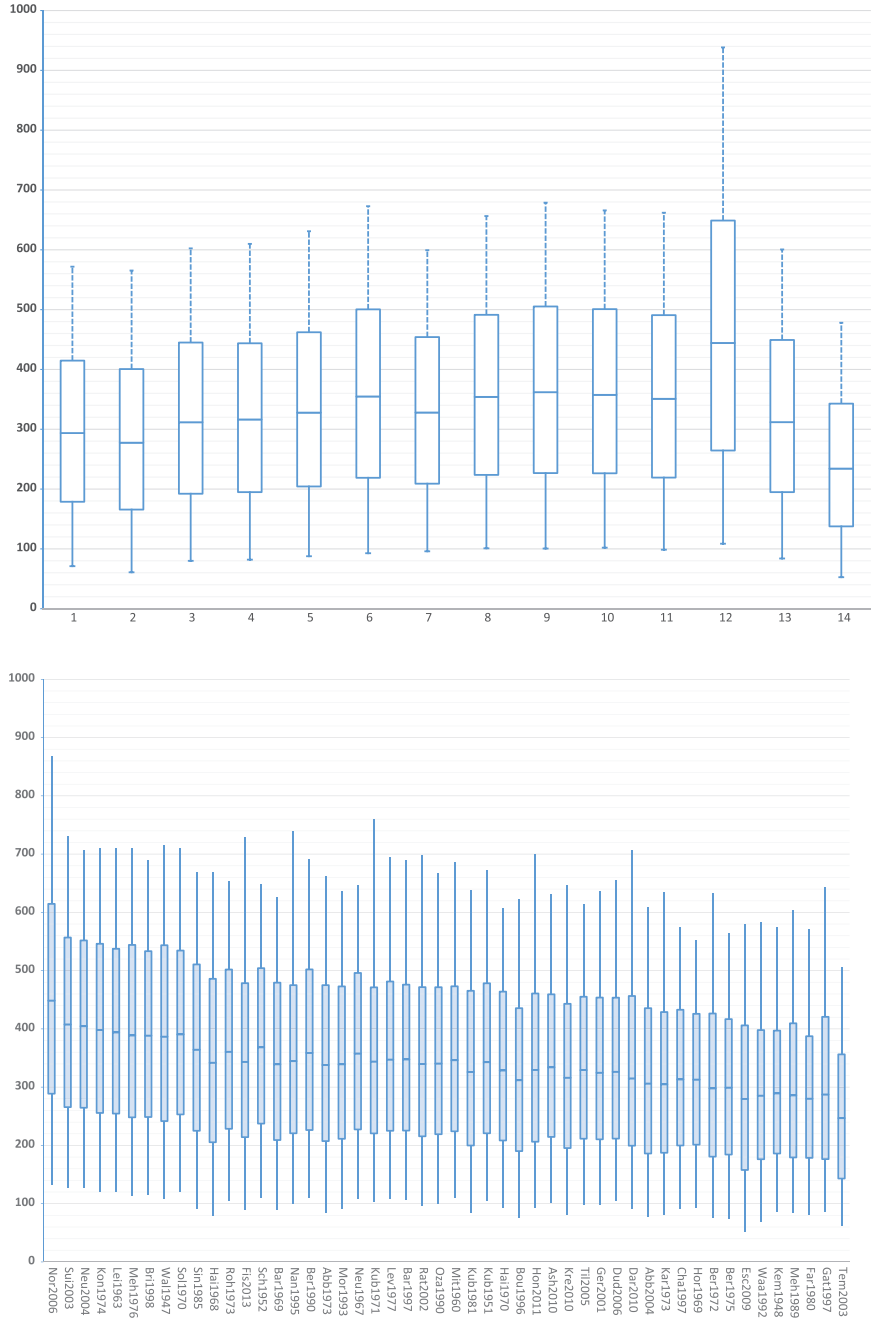

Figure 3.1–2: Box plot of BPM ranges aggregated by measure (top) and by performance (bottom).

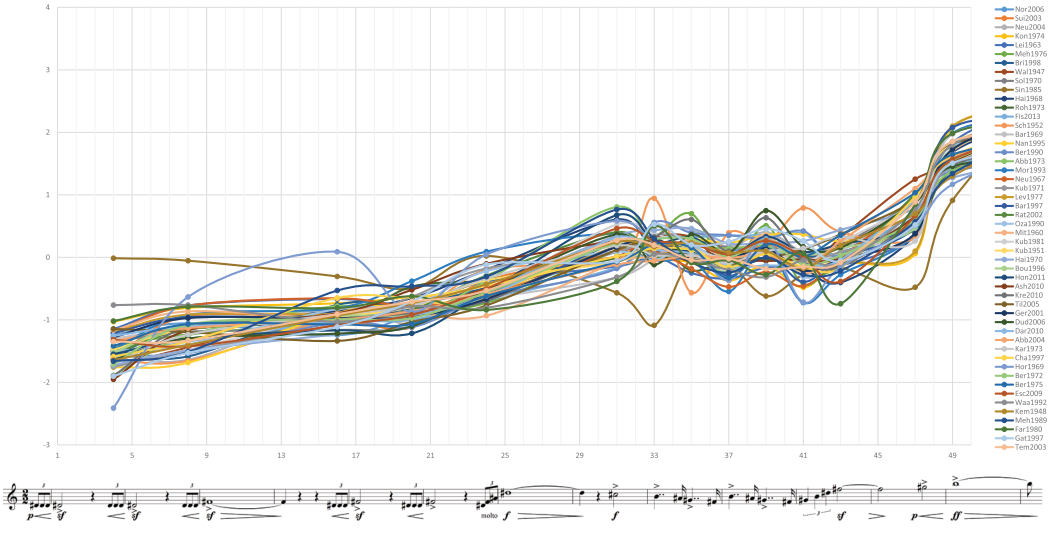

Figure 4: Normalized RMS as a function of music beats. Lines represent RMS variations track by track, and circles are synchronization anchors referring to the score below.

segment, we considered the signal $x_{m,n}$, made of N_m samples. The RMS of the m-th fragment, indicated as ρ_m, was computed through the formula:

$$\rho_m = \sqrt{\frac{1}{N_m} \sum_{n=0}^{N_m-1} x_{m,n}^2}$$

Finally, we normalized RMS values through the following formula:

$$\rho^* = \frac{\rho - \mu(\rho)}{\sigma(\rho)}$$

where $\mu(\rho)$ is the mean and $\sigma(\rho)$ is the standard deviation computed on the RMS values. The resulting diagram, shown in Figure 4, provides a more meaningful comparison about the loudness of original performances, where the baseline represents the average value of normalized RMS for each track. Also in this case, we can notice a similar trend for most of the analysed tracks, with some notable exceptions such as Sin1985 (where the crescendos and diminuendos are often in opposite phase compared to other incipits) and Sch1952 (with several RMS oscillations between beats 31 and 43).

In this section, we have presented two heterogeneous examples of automatic interpretation analysis that may take advantage of a multi-layer format; but the presence of metadata can foster additional types of analysis and new ways to cluster and compare results. For example, our test set contains three performances conducted by Rafael Kubelík, and two of them with the same orchestra, i.e. the Symphonieorchester des Bayerischen Rundfunks. It could be interesting to carry out a more thorough investigation of similarities and differences in interpretation when the same conductor and/or the same orchestra are involved. Unfortunately, this activity would go far beyond the goals of the present paper.

7. Pedagogical implications

The case study presented in the previous section can be generalized, trying to infer a number of pedagogical implications for music teaching and learning.

A first point to highlight is the valence of multi-layer formats per se, since they carry organized and synchronized information, thus providing scholars with a richer view of a learning object. For instance, in the example presented in Section 6, musicologists can take benefit from a comprehensive description of a music piece in its multiple facets, including score symbols, structural information and audio performances (Baratè and Ludovico 2012).

Another educational implication is the possibility to learn by example, thanks to the availability of a potentially high number of instances to be compared to infer both common and distinguishing features (Baratè and Ludovico 2013). In the case of interpretative models, this approach implies the possibility to jump from a performance to another in real time (thanks to intra-layer synchronization), and – if required – to easily refer to the original notation (thanks to inter-layer synchronization).

The concept of *multilayer music-oriented learning object* has been introduced for the first time in (Baratè et al. 2013). The authors started from the commonly accepted definition of *learning object* as a digital file intended to be used for pedagogical purposes, which includes, either internally or via association, suggestions on the appropriate context within which to utilize the object (Sosteric and Hesemeier 2002). In this sense, an IEEE 1599 document can be seen as a learning object that embeds organized information and a number of relationships among content that constitute the context for music learning and experience.

This theoretical approach has been implemented in a number of initiatives. For instance, Ludovico and Mangione (2014) proposed the implementation of an active e-book based on the IEEE 1599 format as a pedagogical tool to foster self-regulation in music education. A live demo is publicly available at http://pearson.lim.di.unimi.it.

The flexibility of the multi-layer approach emerges also from pedagogical applications apparently far from the music domain, such as the implementation of IEEE 1599-based tools for content and language integrated learning in primary school (Ludovico and Zambelli 2017).

A general-purpose viewer for IEEE 1599 documents, including the option to compare different notational instances, to easily jump from a media file to another, to obtain basic statistical data about pitch distribution, note duration, etc., is contained in the *Music Archive* section of the IEEE 1599 portal, available at http://ieee1599.lim.di.unimi.it/ and recently presented to the scientific community (Avanzini et al. 2018).

Multi-layer formats pave the way for many applications in the field of computational musicology, allowing the inference of results from the automatic analysis of aggregated data. In this case, a multi-layer approach does not merely improve the description and comprehension of music, but becomes fundamental to extract information that only a synoptic view can make explicit. For instance, a single document presenting synchronized audio tracks allows an automatic comparison of variations in agogics across different performances.

8. Conclusions and future work

The purpose of this work was to encourage the adoption of multi-layer formats for music description in the context of interpretative modelling, since they facilitate the automatic extraction of useful musical and audio features. Even

if multi-layer formats are not strictly required, as demonstrated by past and current research based on live listening, analogue media or 'traditional' digital formats, they prove to be effective in collecting, organizing and relating huge amounts of data. On one side, this fosters analyses based on multiple, synchronized instances (*1-layer analysis*); on the other, the additional information carried by other layers – such as symbolic, graphical, structural information – improves modelling activities and unveils new possibilities (*multi-layer analysis*).

The results obtained so far thanks to multi-layer description of music can be extended to many domains, including computer-assisted classification, computational musicology, content-driven recommender systems and re-synthesis of interpretative models.

Currently, the main research problem to tackle concerns the production of multi-layer materials, which – in absence of reliable algorithms for automatic recognition, feature extraction and synchronization – is a time-consuming operation and requires the supervision of experts. Solving such an issue is one of the directions our research will take in the near future. In this sense, an advantage offered by multi-layer formats is the possibility to take benefit from already available information to conduct an 'informed' analysis of new content. For example, the optical music recognition of a new score version has not to infer score symbols from scratch, rather it has to correctly locate the position of already-known symbols over the new page, starting from information contained in the logic and, potentially, in the notational layers.

Another field to be further investigated regards the potential of artificial-intelligence techniques (e.g., deep learning) to infer performance information from a multi-layer environment. This aspect has been explored in a number of contributions presented at the *1st International Workshop on Multilayer Music Representation and Processing*.

Concerning the multi-layer format that we adopted, namely IEEE 1599, the working group who originally developed it has been recently reconstituted, thus a new version of the standard is expected to be released in a few years. It could be the occasion to improve those parts devoted to computational musicology, such as the structural layer, and to integrate new formats and descriptional approaches.

REFERENCES

Arbo, A. (2015), 'Music and technical reproducibility: A paradigm shift', in G. Borio (ed.), *Musical Listening in the Age of Technological Reproduction*, Abingdon-on-Thames: Routledge, pp. 53–67.

Atkinson, R. K. and Renkl, A. (2007), 'Interactive example-based learning environments: Using interactive elements to encourage effective processing of worked examples', *Educational Psychology Review*, 19:3, pp. 375–86.

Avanzini, F., Baratè, A., Haus, G., Ludovico, L. A., Mauro, D. A., Ntalampiras, S. and Presti, G. (2018), 'Quale futuro per il formato IEEE 1599?', in F. Fontana and A. Gulli, A. (eds), *Machine Sounds, Sound Machines. Atti del XXII CIM - Colloquio di Informatica Musicale*, XXII, Venezia: DADI - Dip. Arti e Design Industriale, Università IUAV di Venezia, pp. 115–21.

Baggi, D. L. and Haus, G. M. (2013), *Music Navigation with Symbols and Layers: Toward Content Browsing with IEEE 1599 XML Encoding*, Hoboken: John Wiley and Sons.

Baratè, A., Bergomi, M. and Ludovico, L. A. (2013), 'Development of serious games for music education', *Je-LKS: Journal of e-Learning and Knowledge Society*, 9:2, pp. 93–108.

Baratè, A. and Ludovico, L. A. (2012), 'New frontiers in music education through the IEEE 1599 standard', *International Conference on Computer Supported Education*, Setúbal: SciTePress, pp. 146–51.

——— (2013), 'IEEE 1599 applications for entertainment and education', in D. Baggi and G. Haus (eds), *Music Navigation with Symbols and Layers: Toward Content Browsing with IEEE 1599 XML Encoding*, Hoboken: John Wiley and Sons, pp. 115–32.

Bellini, P., Nesi, P., Campanai, M. and Zoia, G. (2006), *ISO/IEC FCD 14496-23:200x - Symbolic Music Representation*.

Bellini, P., Nesi, P. and Zoia, G. (2005), 'Symbolic music representation in MPEG', *IEEE MultiMedia*, 12:4, pp. 42–49.

Cairncross, S. and Mannion, M. (2001), 'Interactive multimedia and learning: Realizing the benefits', *Innovations in Education and Teaching International*, 38:2, pp. 156–64.

Clarke, E. F. (1988), 'Generative principles in music performance', in E. F. Clarke and J. A. Sloboda (eds), *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, New York: Clarendon Press and Oxford University Press, pp. 1–26.

Cook, N. (2007), 'Performance analysis and Chopin's mazurkas', *Musicae scientiae*, 11:2, pp. 183–207.

D'Aguanno, A. and Vercellesi, G. (2007), 'Automatic synchronisation between audio and score musical description layers', in B. Falcidieno, M. Spagnuolo, Y. Avrithis, I. Kompatsiaris and P. Buitelaar (eds), *Semantic Multimedia. SAMT 2007. Lecture Notes in Computer Science*, 4816, Berlin, Heidelberg: Springer, pp. 200–10.

Damm, D., Fremerey, C., Thomas, V., Clausen, M., Kurth, F. and Müller, M. (2012), 'A digital library framework for heterogeneous music collections: From document acquisition to cross-modal interaction', *International Journal on Digital Libraries*, 12:2&3, pp. 53–71.

Faiella, F. and Mangione, G. R. (2012), *E-Learning: le pratiche consolidate e i nuovi scenari di ricerca*, San Cesario di Lecce: Pensa Editore.

Friberg, A. and Sundströöm, A. (2002), 'Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern', *Music Perception: An Interdisciplinary Journal*, 19:3, pp. 333–49.

Gabrielsson, A. (2003), 'Music performance research at the millennium', *Psychology of Music*, 31:3, pp. 221–72.

Goebl, W., Pampalk, E. and Widmer, G. (2004), 'Exploring expressive performance trajectories: Six famous pianists play six Chopin pieces', Proceedings of the 8th International Conference on Music Perception and Cognition, Sydney: Causal Productions, pp. 505–09.

Gog, T. Van and Paas, F. (2008), 'Instructional efficiency: Revisiting the original construct in educational research', *Educational Psychologist*, 43:1, pp. 16–26.

Good, M. and Actor, G. (2003), 'Using MusicXML for file interchange', Proceedings of the Third International Conference on Web Delivering of Music, 2003 (WEDELMUSIC 2003), New York: IEEE, p. 153.

Green, L. (2017), *How Popular Musicians Learn: A Way Ahead for Music Education*, Abingdon-on-Thames: Routledge.

Haus, G. and Longari, M. (2005), 'A multi-layered, time-based music description approach based on XML', *Computer Music Journal*, 29:1, pp. 70–85.

Haus, G. and Ludovico, L. A. (2006), 'The digital opera house: An architecture for multimedia databases', *Journal of Cultural Heritage*, 7:2, pp. 92–97.

Liem, C., Müller, M., Eck, D., Tzanetakis, G. and Hanjalic, A. (2011), 'The need for music information retrieval with user-centered and multimodal strategies', Proceedings of the 1st International ACM Workshop on Music Information Retrieval with User-Centered and Multimodal Strategies, New York: ACM, pp. 1–6.

Lindsay, A. and Kriechbaum, W. (1999), 'There's more than one way to hear it: Multiple representations of music in MPEG-7', *Journal of New Music Research*, 28:4, pp. 364–72.

Ludovico, L. A. and Mangione, G. R. (2014), 'An active e-book to foster self-regulation in music education', *Interactive Technology and Smart Education*, 11:4, pp. 254–69.

Ludovico, L. A. and Zambelli, C. (2017), 'Web-based frameworks for CLIL in primary school: Design, implementation, pilot experimentation and results', in G. Costagliola, B. M. McLaren, J. Uhomoibhi and S. Zvacek (eds), *Computers Supported Education - 8th International Conference, CSEDU 2016, Rome, Italy, April 21-23, 2016, Revised Selected Papers, Communications in Computer and Information Science*, vol. 739, Berlin, Heidelberg: Springer, pp. 139–58.

Mayer, R. E. (2002), 'Multimedia learning', *Psychology of Learning and Motivation*, 41, pp. 85–139.

Mazzola, G. and Göller, S. (2002), 'Performance and interpretation', *Journal of New Music Research*, 31:3, pp. 221–32.

Merrienboer, J. J. Van and Sweller, J. (2005), 'Cognitive load theory and complex learning: Recent developments and future directions', *Educational Psychology Review*, 17:2, pp. 147–77.

Moreno, R. and Mayer, R. (2007), 'Interactive multimodal learning environments', *Educational Psychology Review*, 19:3, pp. 309–26.

Najjar, L. J. (1996), 'Multimedia information and learning', *Journal of Educational Multimedia and Hypermedia*, 5:2, pp. 129–50.

Paas, F. G. and Van Merriënboer, J. J. (1993), 'The efficiency of instructional conditions: An approach to combine mental effort and performance measures', *Human Factors*, 35:4, pp. 737–43.

Palmer, C. (1996), 'Anatomy of a performance: Sources of musical expression', *Music Perception: An Interdisciplinary Journal*, 13:3, pp. 433–53.

Peeters, G., Giordano, B. L., Susini, P., Misdariis, N. and McAdams, S. (2011), 'The timbre toolbox: Extracting audio descriptors from musical signals', *The Journal of the Acoustical Society of America*, 130:5, pp. 2902–16.

Priest, P. (1989), 'Playing by ear: Its nature and application to instrumental learning', *British Journal of Music Education*, 6:2, pp. 173–91.

Repp, B. H. (1990), 'Patterns of expressive timing in performances of a Beethoven minuet by nineteen famous pianists', *The Journal of the Acoustical Society of America*, 88:2, pp. 622–41.

——— (1992), 'Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's *Träumerei*', *The Journal of the Acoustical Society of America*, 92:5, pp. 2546–68.

Roland, P. (2002), 'The music encoding initiative (MEI)', Proceedings of the First International Conference on Musical Applications Using XML (MAX 2002), New York: IEEE, pp. 55–59.

Sankey, M., Birch, D. and Gardiner, M. (2010), 'Engaging students through multimodal learning environments: The journey continues', Proceedings ASCILITE

2010: 27th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Curriculum, Technology and Transformation for an Unknown Future, Brisbane: University of Queensland, pp. 852–63.
Scheirer, E. D. (1998), 'The MPEG-4 structured audio standard', Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, New York: IEEE, pp. 3801–04.
Shaffer, L. H. and Todd, N. P. (1987), 'The interpretive component in musical performance', *Action and Perception in Rhythm and Music*, Stockholm: R. Swed. Acad. Music, pp. 139–52.
Sosteric, M. and Hesemeier, S. (2002), 'When is a learning object not an object: A first step towards a theory of learning objects', *The International Review of Research in Open and Distributed Learning*, 3:2, pp. 1–16.
Steyn, J. (2002), 'Framework for a music markup language', Proceeding of the First International IEEE Conference on Musical Application using XML (MAX 2002), New York: IEEE, pp. 22–29.
Sweller, J., Ayres, P. and Kalyuga, S. (2011), *Cognitive Load Theory*, Berlin, Heidelberg: Springer Science+Business Media.
Sweller, J., Merrienboer, J. J. Van and Paas, F. G. (1998), 'Cognitive architecture and instructional design', *Educational Psychology Review*, 10:3, pp. 251–96.
Vickers, E. (2010), 'The loudness war: Background, speculation, and recommendations', *Audio Engineering Society Convention 129*, November 2010, New York: Audio Engineering Society.
Widmer, G. and Goebl, W. (2004), 'Computational models of expressive music performance: The state of the art', *Journal of New Music Research*, 33:3, pp. 203–16.
Woody, R. H. (2012), 'Playing by ear: Foundation or frill?', *Music Educators Journal*, 99:2, pp. 82–88.
Wrightson, K. (2000), 'An introduction to acoustic ecology', *Soundscape: The Journal of Acoustic Ecology*, 1:1, pp. 10–13.
Yue-guo, G. U. (2007), 'On multimedia learning and multimodal learning', *Computer-Assisted Foreign Language Education*, 2:1.

SUGGESTED CITATION

Baratè, A., Haus, G., Ludovico, L. A. and Presti, G. (2019), 'Investigating interpretative models in music through multi-layer representation formats', *Journal of Music, Technology & Education*, 12:1, pp. 95–113, doi: 10.1386/jmte.12.1.95_1

CONTRIBUTOR DETAILS

Adriano Baratè is a researcher at the Laboratorio di Informatica Musicale (LIM), Department of Computer Science, University of Milan. His research interests include Music Petri Nets, formalization and encoding of symbolic music, cultural heritage and computational musicology. Baratè has a Ph.D. in computer science from the University of Milan. He was a member of the IEEE Technical Committee on computer generated music and is currently part of the W3C Music Notation Community Group and the IEEE Working Group for XML Musical Application.

E-mail: adriano.barate@unimi.it

https://orcid.org/0000-0001-8435-8373

Goffredo Haus is a full professor of Computer Science at the University of Milan. He has a master's degree in physics from the University of Milan. He founded the Laboratorio di Informatica Musicale (LIM) in 1985 and has the bachelor's degree in music information science in 2001. His research interests include multimedia and human–computer interaction for music and cultural heritage. He was official reporter of the IEEE 1599 standard, chair of the IEEE Technical Committee on Computer Generated Music, and he is currently chair of the IEEE Working Group for XML Musical Application.

E-mail: goffredo.haus@unimi.it

https://orcid.org/0000-0002-3477-4042

Luca A. Ludovico is a researcher at the Laboratorio di Informatica Musicale (LIM), Department of Computer Science, University of Milan. He graduated in computer engineering at the Politecnico of Milan and obtained a Ph.D. in computer science at the University of Milan. His main research activity consists in the investigation of computer-based encoding for symbolic aspects of music. As a member of the IEEE Technical Committee on Computer Generated Music, he has been the Italian coordinator of the project PAR1599 that brought to the standardization of the IEEE 1599 format. Currently, he is the vice-chair of the IEEE Working Group for XML Musical Application and a member of the MIDI Association and of the W3C Music Notation Community Group.

Contact: Dipartimento di Informatica Giovanni Degli Antoni, Università degli Studi di Milano, via G. Celoria 18, 20133 Milano, Italy.
E-mail: luca.ludovico@unimi.it

https://orcid.org/0000-0002-8251-2231

Giorgio Presti received the bachelor, master and Ph.D. degrees in computer science from the University of Milan in 2009, 2013 and 2017, respectively. Currently he is a postdoctoral researcher at the Laboratorio di Informatica Musicale (LIM) of the same university. His research interests include signal processing, sound and music computing, affective computing, sonification and scientific dissemination, especially by means of art.

E-mail: giorgio.presti@unimi.it

https://orcid.org/0000-0001-7643-9915