Information Systems 73 (2018) 35–49

Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/is

Modeling Music as Synchronized Time Series: Application to Music Score Collections

Raphaël Fournier-S'niehotta, Philippe Rigaux*, Nicolas Travers

Conservatoire National des Arts et Métiers (CNAM), France

ARTICLE INFO

Article history:
Received 21 June 2017
Revised 15 October 2017
Accepted 10 December 2017
Available online 16 December 2017

ABSTRACT

Music is a temporal organization of sounds, and we can therefore assume that any music representation has a structure that reflects some conceptual principles. This structure is hardly explicitly accessible in many encodings, such as, for instance, audio files. However, it appears much more clearly in the language of *music notation*.

We propose to use the music notation language as a framework to model and manipulate the content of digital music objects, whatever their specific encoding may be. We describe an algebra that relies on this structured music representation to extract, restructure, and search such objects. The data model leverages the hidden structure of digital music encodings to enable powerful manipulations of their content.

We apply the model to collections of music scores. We describe a system, based on an extension of XQuery with our algebra, that provides search, reorganization, and extraction functionalities on top of large collections of XML-encoded digital scores. Beyond its application to music objects, our work shows how one can rely on a structured content embedded in a complex XML encoding to develop robust collection management tools with minimal implementation effort.

© 2017 Published by Elsevier Ltd.

1. Introduction

Music is a temporal organization of sounds, and we can therefore assume that music content has a structure that reflects some conceptual and organisational principles. Digital encoding of music is mostly represented by audio files, in which this structure is blurred and difficult to capture accurately. But music content can also be encoded in a notational form that has been used for centuries to preserve and exchange music works. Music notation appears to be a mature language to represent the discrete, typed, coordinated elements that together constitute the description of complex music pieces.

The structure of music notation

Fig. 1 shows a first example of a *monophonic score*, the famous "Ode to joy" theme from Beethoven's 9th symphony. The central element is a *note* representing the intended production of a single sound. A note symbol is a black or white dot that carries two essential informations: the *frequency* of the intended sound, measured in Hertz (Hz), and its *duration*, which is a value relative to the *beat*.

Encoding frequencies. The hearing frequency of human beings ranges approximately from 20 to 20,000 Hz. Although this range is in principle continuous, most music compositions rely on a discretization in two steps. First, the frequency range is partitioned in *octaves*: a note *a* is one octave above another note *b* if the frequency of *a* doubles that of *b*. Octaves are in turn divided in twelve equal *semi-tones*. Since the usual frequency range covers 8 octaves, one obtains $8 * 12 = 96$ possible frequency levels, or *pitches*.

These levels are materialized by horizontal lines on a score. A note can be positionned on a line or between two lines. The whole grid, covering the whole frequency range, would display $96/2 = 48$ lines and take a lot of useless space. Since the usual range of a single musician or singer is much more limited, music scores uses a 5-lines grid, called a *staff*, whose relative position in the complete grid is given by an initial symbol, the *clef* (Fig. 1).

Staves and clefs are example of semiotic artefacts mostly irrelevant in the perspective of using music notation as a language encoding music content. If we leave apart layout concerns, there exists a simple, non graphic convention to encode a music note. Inside an octave, seven pitches are related by close mathematical relationships, and form the Pythagorean scale (commonly called *diatonic scale* nowadays). They are encoded by a letter (A, B, C, ... G).

* Corresponding author.
E-mail addresses: Raphael.fournierSniehotta@cnam.fr (R. Fournier-S'niehotta), Philippe.rigaux@cnam.fr (P. Rigaux), Nicolas.travers@cnam.fr (N. Travers).

https://doi.org/10.1016/j.is.2017.12.003
0306-4379/© 2017 Published by Elsevier Ltd.

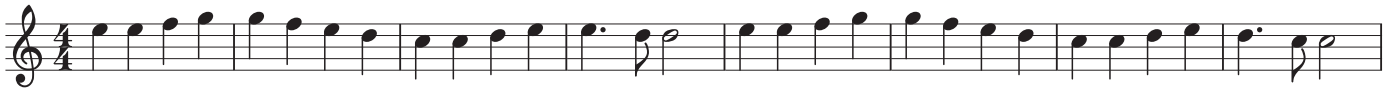

Fig. 1. Beethoven's *Ode to Joy* theme.

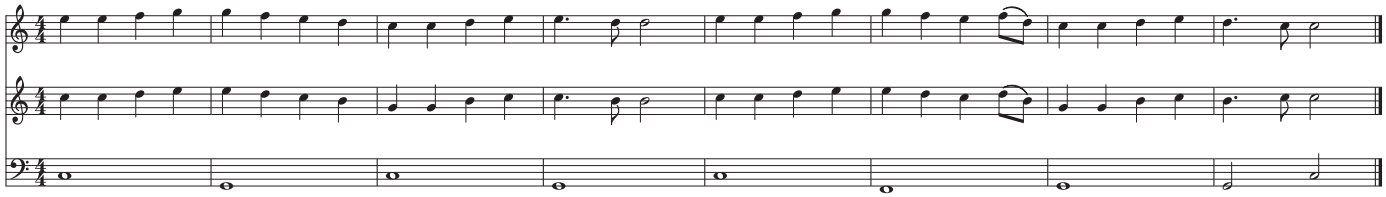

Fig. 2. *Ode to Joy*, three instruments.

The other pitches can be obtained by either adding (symbol ♯) of substracting (symbol ♭) a semi-tone. To summarize: any pitch in common music notation is encoded by (i) a diatonic letter, (ii) an octave in the range 1-8, and (optionally) (iii) a ♯ or a ♭ to denote a semi-tone up/down the diatonic frequency. The score of Fig. 1 starts with two E5 (frequency 659.25 Hz), followed by two F5 (frequency 698.46 Hz), two G5 (783.99 Hz), etc.

Encoding beats and durations. Time is discrete, and temporal values are expressed with respect to the *beat*, a pulse that (in principle) remains constant throughough a same piece of music. The *time signature*, a rational (here, 4/4) gives the beat unit, and the temporal organization of the music in groups of beats, or *measures*. In our example, the denominator states that the beat corresponds to a black note (a quarter of the maximal note duration), and the numerator means that each measure contains four beats. All the possible durations are obtained by applying a simple ratio to the beat: a white note is twice a quarter, a hamped black is half a quarter, etc.

From these (basic) explanations, it follows that music notation can be seen as a way to represent sounds in a 2-dimensional space where each axis (frequencies and durations) is discretized according to some simple rules based on proportional relationships. Moreover, there exists a simple and commonly used encoding to denote each point in this space. We will use this discrete sound domain as a basis of our data model.

Scores as times series. Let us now turn our attention to the *sequence* of notes in Fig. 1. An implicit constraint is that a note starts immediatly after the end of its predecessor. In other words, there is no overlapping of the timespans covered by two distinct notes. This is natural if we consider that the original intent of this notation is to encode the music part assigned to a single singer, who can hardly produce simultaneous sounds. This part is accordingly called a *voice*, and we will use this term to denote the basic structure of music objects representation as time series of musical events, assigned to timespans that do not overlap with one another.

Polyphonic scores are represented as a combination of several voices. Fig. 2 gives an illustration (the same theme, excerpt of the orchestra parts). In terms of music content, the important information expressed by the notation is the *synchronization* of sounds, graphically expressed by their vertical alignment. The first sound for instance is an harmonic combination of three notes: a C3, a C5, and a E5 (from bottom to top). The two upper notes share the same duration (a quarter), but the bottom one (the bass) is a whole note. This single note is therefore synchronized with 4 notes of the two upper parts.

In such complex scores, one can always adopt two perspectives on the music structure. A vertical perspective, called *harmonic*, focuses on the vertical superposition of sounds, whereas a horizontal one, called *polyphonic*, rather considers the sequential development of each voice. Those two aspects constitute, beyond all the semiotic decorations related to the graphic layout of a score, what could be called the "semantic" of music notation, since they encode all information pertaining to the sounds and their temporal organization (at least for the part of this information conveyed by the notation; some other important features, such as intensity and timbre, are left to the performer's choice). Together, they define what we will consider in the following as the (structured) *music content*.

Music pieces as synchronized time series

Finally, this modeling perspective can be extended to cover the more general concept of synchronized time series built from arbitrary value domains. Consider our third example shown on Fig. 3, the same *Ode to Joy* enriched with lyrics. The lower staff consists of a single voice, the bass. The upper one is a vocal part which, in our model, consists of two voices, the first one composed of sounds, and the second one of syllables. The latter is an example of a temporal function that, instead of mapping timestamps to sounds, maps timestamps to syllables.

Position, goals and contributions

The position adopted in the present paper relies on two ideas. Firstly, music notation is a proven, sophisticated, powerful formal language that provides the basis of a data model for music content. Secondly, instances of this model can be extracted from digital music documents, and this extraction yields a structured representation of this object through which its content can be inspected, decomposed, transformed and combined with other contents. In the context of large collections of such music objects, this opens perspectives for advanced search, indexing and data manipulation mechanisms.

We can therefore envision a modeling that abstracts the music content as a synchronization (harmonic view, expressed by the vertical axis in the score representation) of temporal sequences of acoustic events (polyphonic view, expressed by the horizontal axis). As a natural generalization of this modeling approach, we accept polymorphic events that can either represent sounds, or features that make sense as time-dependent information synchronized with the music content. Fig. 4 summarizes the envisioned system. The bottom layer is a Digital Music Library (DML) managing music objects in some encoding, whether audio (WAV, MP3, MIDI), image (PDF, PNG), or XML (MusicXML, MEI). Such encodings are not designed to support content-based manipulations, and, as a matter of fact, it is hardly possible to do so. However, we can *map* the encoding toward a model layer where the content is extracted and structured according to the model structures. This automatic mapping is called *transcription* for audio files, *optical music recognition* (OMR) for images, and is a much simpler extraction for

Download English Version:

https://daneshyari.com/en/article/6858620

Download Persian Version:

https://daneshyari.com/article/6858620

Daneshyari.com