THE SEGMENT ONTOLOGY: BRIDGING MUSIC-GENERIC AND DOMAIN-SPECIFIC

Ben Fields *, *Tim Crawford*

Department of Computing
Goldsmiths, University of London
contact: `b.fields@ieee.org`

Kevin Page, David De Roure

Oxford eResearch Centre
University of Oxford

ABSTRACT

Existing semantic representations of music analysis encapsulate narrow sub-domain concepts and are frequently scoped by the context of a particular MIR task. Segmentation is a crucial abstraction in the investigation of phenomena which unfold over time; we present a Segment Ontology as the backbone of an approach that models properties from the musicological domain independently from MIR implementations and their signal processing foundations, whilst maintaining an accurate and complete description of the relationships that link them. This framework provides two principal advantages which are explored through several examples: a layered separation of concerns that aligns the model with the needs of the users and systems that consume and produce the data; and the ability to link multiple analyses of differing types through transforms to and from the Segment axis.

Index Terms— Music Analysis, Data Modeling, RDF-OWL, Segmentation, Structural Analysis

1. INTRODUCTION AND BACKGROUND

Music information retrieval (MIR) has, as is the case with many other data-centric domains of study, spent the last several years on a march toward ever larger datasets. This is seen recently in a number of projects and shared datasets (e.g. Structural Analysis of Large Amounts of Music Information, SALAMI [1] and the Million Song Dataset[1]). As the quantity of data continues to grow, many potential research questions can be envisaged based on the comparison and combination of large quantities of MIR algorithmic output; to support use (and re-use) of data in this way attention must be paid to the way it is stored, modeled and published. It has already been shown that using a Linked Data approach can enable joins of this nature at the level of signal and collections [2], though the prototype presented in that work was focused on one analysis (genre) and therefore only on results from that task. In an effort to model the MIR task itself in more detail, and enable Linked Data joins at the result level, we present the Segmentation Ontology, focused on modeling division of temporal-signal (principally music) into subunits.

The remainder of this paper will detail the ontology: after introducing the conceptual framework upon which the ontology is based and existing complementary ontologies used in our approach, we detail the classes and properties used; we then devote a significant portion of the paper to worked examples, and compare our information representation with others currently in use.

---

*B. Fields is also affiliated with the School of Electronics and Computer Science at the University of Southampton
[1]`http://labrosa.ee.columbia.edu/millionsong/`

2. FOUNDATIONAL CONCEPTS

Many systems developed for MIR tasks are constructed of common elements. To support the "joining" of disparate MIR components into a complete system, and to enable the use of analytic output by domain experts (e.g. musicologists) we consider the concepts core to each, and broadly categorize these as:

1. **Domain-specific** musicology: concepts, in our use case, from musicology, and the human interpretation of music and sound.
2. **Domain-specific** MIR tasks: parts of the model that relate to an MIR task, such as the elements extracted by a feature extractor, common labels from a classifier, distance metrics from a system such as [3].
3. **Music-generic**: common concepts that transcend domain-specific as Intervals, Segments, etc.
4. **High-level Relationships**: the absolute and relative relationships between music-generic elements, TimeLines and SegmentLines; and the maps between them.

To illustrate this conceptual distinction we consider an example of structural segmentation:

1. **Domain-specific** musicology concepts are elements of form, such as intro, verse, chorus, bridge; sonata, minuet and trio, fugue. These are likely to be applied to sections of the signal, for example "this section is a bridge".
2. **Domain-specific** MIR tasks encompass artifacts of the structural segmentation task, for example a classifier might identify (and potentially label) sections that are similar; a contributor task might identify chords. Again, these concepts are likely to be applied to sections of signal.
3. **Music-generic** concepts are common to different tasks and applications. Here the segments would be those annotated using the domain-specific concepts and the alignments and relationships between them (e.g. the segment labelled as a chorus follows the segment labelled as a verse; that one chord follows another).
4. Finally **high-level relationships** capture mappings between the musicologically labelled segments and the MIR task derived segments. We expand upon this example in Figure 6.

A further requirement when considering MIR tasks is the ability to capture provenance of both data and method: for example the algorithmic elements used by the tasks including the software versions and how and when they were run; or identifying factors of human-generated ground truth.

3. RELATED MODELS

A number of existing ontologies are relevant and either extended by or used in conjunction with the Segment Ontology.

The Timeline Ontology (TL) primarily describes discrete temporal relationships. Following early development for the signal processing domain it has been more widely used to describe temporal placement and alignment of Things [4]. It also introduces the TimeLineMaps classes which encode an explicit mapping from one TimeLine to another, e.g. from a ContinuousTimeLine to a DiscreteTimeLine via a UniformSamplingMap. It explicitly names AbstractTimeLines but, to the authors knowledge, no examples using this and the associated Maps exist or are in use. The TimeLine ontology is used directly or through alignment with equivalent relative concepts throughout our approach and our examples.

The Music Ontology (MO) models high-level concepts about and around music including editorial, cultural and acoustic information [5]. To express temporal information it incorporates both the Timeline and Event Ontologies. We link to the Music Ontology through instances of audio signal against which we are asserting segmentation and domain-specific labeling.

The Similarity Ontology (SIM), sometimes referred to as MuSim, was conceived to model music similarity [6]. The current version's use of b-nodes to express associations between class instances allows an efficient general unnamed representation of any type of association (so the ontology could perhaps be more aptly described as one for "associations"). We use the Similarity Ontology throughout our approach to associate music-generic and domain-specific concepts.

4. ONTOLOGY AND APPROACH

While the Segment Ontology that follows is the backbone of our approach, it is only a mechanism to facilitate our overall method: recognizing that there can, and should, be many models of domain specific knowledge, and that music-generic and high-level relationships be used to move across these boundaries and make links between the knowledge within. As such, we use Segments as a music-generic dimension between explicitly temporal and implicitly (or indirectly) temporal concepts (and ontologies).

The core concepts and properties in the Segment Ontology are shown in Figure 1 and detailed below:

- **Segment**: an Interval with a "label" expressing an association (SIM) that can be "placed" upon TimeLines (TL) and SegmentLines. There are five intra-Segment properties, to express alignment or membership: **segmentBefore**, **segmentAfter**, **segmentBegins**, **segmentEnds**, and **contains**. These are all sub-properties from TL with the exception of contains, a property necessary when alignment or membership cannot be inferred from time (e.g. from NonSequentialMap).
- **SegmentLine**: a relative complement to TimeLine
- **SegmentLineMap**: a means to express a high-level relationship between SegmentLines or with TimeLines; can imply relationships between Segments on SegmentLines and TimeLines; similarly a SegmentLineMap can be used to infer properties between Segments. Three subclasses are specified: **RatioMap** a fixed integer number of Segments mapped from one SegmentLine to another; **NonLinearMap**, mapping is not fixed across SegmentLines, however sequential order of Segments is preserved; and **NonSequentialMap**, the least spec-

ified, whereby sequential order of Segments is not preserved across SegmentLines.

While MIR-task and musicology annotations are likely to be described independently, as described in section 2, we model the relationships that stem from these as domain-specific terms in the same way: as (associative) annotations to Segments, Segment- and TimeLines, and the high-level relationships between them.
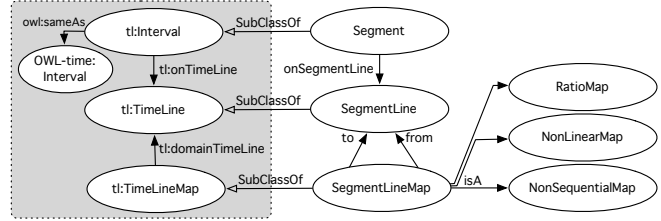

Fig. 1. The class structure of the Segment Ontology. Concepts from TL are on the grey background.

5. WORKED EXAMPLES

Throughout these examples we reference and compare an existing analysis of the Beatles "Help!"[2]. Figure 2 is a generic visualization of the analytic structures that can be found in this piece of music; it is worth recognizing that although Figure 2 does not use any specific ontology or data structure, it does invoke a temporal dimension most would apply as their default interpretation.

Fig. 2. Segmentation of the song "Help!" by The Beatles by song structure, chord, and beat, with alignment shown.

Throughout the examples we have also arranged the models according to the categorization introduced in section 2 to demonstrate how the Segment Ontology enables an approach that bridges these concepts, that is: "R" for High-level Relationships, "M" for music-generic, and "D" for domain-specific.

We also introduce the notion of a "Mythical Music Taxonomy", which represents an ontological structure describing musicological knowledge (as distinct from MIR domain-specific), and the detail of which is beyond the scope of this paper.

Figure 3 shows **structural segmentation with a discrete TimeLine**. The analysis is a ground truth, performed by a human (captured using sim:method), and the relationship between the ground truth label (e.g. "Verse") and the segment is through a b-node from the Similarity Ontology. Segments are tied to a physical TimeLine, and the sequencing of Segments is through explicit temporal markers (times) on that TimeLine. The relationship between the artistic

---

[2] http://isophonics.net/files/annotations/all/ The\%20Beatles/05_-_Help!/01_-_Help!.ttl
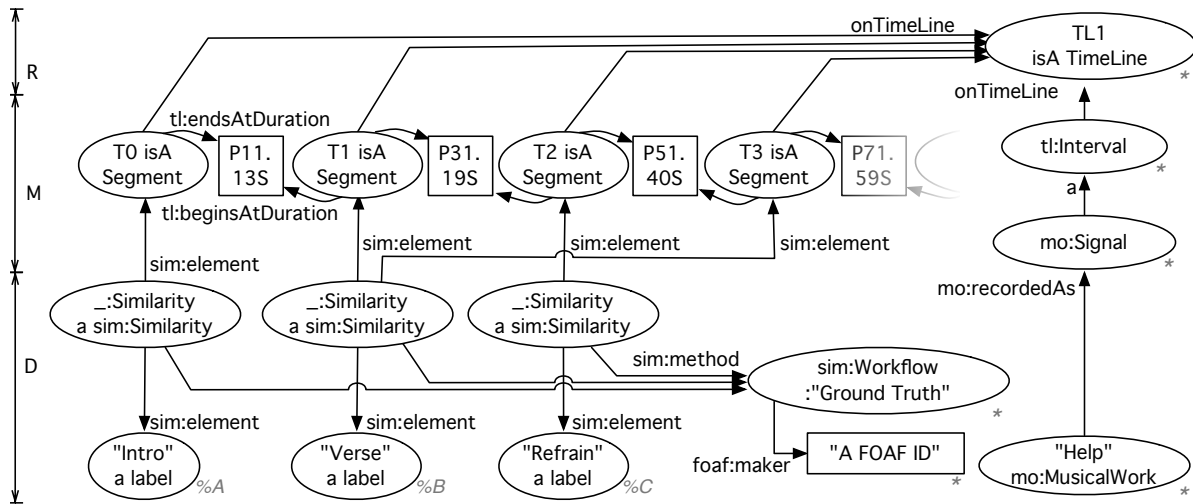
Fig. 3. Structural segmentation modeled with a discrete TimeLine.

work ("Help!") and the analysis is through a recording (a Signal) that is also tied to the TimeLine; this representation is also used in the subsequent examples.

Figure 4 shows **structural segmentation with a relative SegmentLine**, the result of using text analysis of lyrics to perform (relative) structural segmentation. Again the procedure (in this case an algorithm) is shown as sim:method as in Figure 3. Note that the segments are just given a label, e.g. "Verse", or "A", but with no meaning.

Figure 6 gives an example of **mapping a TimeLine to a SegmentLine** using a RatioMap to transform a SegmentLine against a physical timeline. One sequence of Segments is set on a TimeLine and the other sequence of Segments is set against a SegmentLine. The RatioMap enables us to infer the TimeLine from the SegementLine (or vice-versa) but is is optional: the same relationship could be made explicitly.

Figure 5 relates segmented analysis to musical concepts, an extension of Figure 4 into the musicological domain. In addition to the simple labels typically used for classification by a machine learning algorithm, here we can also represent classification of a Segment to the specific verse of this specific work, and the relationship from that specific verse to the musicological concept of "Verse" (as represented in the Mythical Music Taxonomy). Note that this could have been achieved through the alignment of two different analyses (e.g. Figures 3 and 4) via a mapping (e.g. Figure 6).

Figure 7, a **genre analysis with weighting**, shows how frame-by-frame genre analysis from an MIR system output and a frame-derived timeline is represented in our model. Here the genre concepts come from Mythical Music Taxonomy; similarity assertions go directly to genre concepts from the Mythical Music Taxonomy; sim:method is an MIR genre analysis algorithm.

Figure 8 shows the **modeling of domain concepts**, specifically beat and time signatures. Using the Segment structures to model domain concepts as well as tasks allows us to join and map them across analyses (see also Figures 7 and 6). In this case ratio mapping is an instantiation of a time signature.

6. COMPARISON WITH EXISTING REPRESENTATIONS

6.1. Audio Features Ontology

The Audio Feature Ontology (AF) was added to the Sonic computational analysis suite [7] as the principal method of exchanging data for input and output by algorithms within the tools (in conjunction with the Music Ontology). It represents concepts associated with different tasks by explicitly sub-classing the TimeLine Interval, e.g. the StructuralSegment, the ChordSegment and the StructuralSegment classes [8]. We contend that this representation is strongly task-specific, and by conflating domain-specific and music-generic concepts, presents significant barriers to merging and comparing with and between different analysis tasks and musicological knowledge structures.

For example, an encoding of a structural analysis of "Help!" using AF, comparable to the example in Figure 3 contains this RDF graph (serialized in ttl here):

```
:chord_000002
    event:time [
        tl:beginsAt "PT1.053S"^^xsd:duration ;
        tl:duration "PT2.541S"^^xsd:duration ;
        tl:onTimeLine :timeline_2669d37605ce ;
        a tl:Interval
    ] ;
    a af:ChordSegment ;
    rdfs:label "B:min" .
:chord_000003
    event:time [
        tl:beginsAt "PT3.594S"^^xsd:duration ;
        tl:duration "PT2.496S"^^xsd:duration ;
        tl:onTimeLine :timeline_2669d37605ce ;
        a tl:Interval
    ] ;
    a af:ChordSegment ;
    rdfs:label "G" .
```

The task-specific subclassing in AF (here af:ChordSegment) makes the joins and mappings shown in Figures 4, 6, and 5 much more complicated, and in some cases impossible, to express in AF. The ability to do so is a direct consequence of the Segment Ontology's explicit separation of domain-specific and music-generic, which simplifies these joins and mappings.

6.2. MIREX

Although we believe the Segment Ontology can enable new methods for joining data and algorithms, it is also important that supports existing use cases in the community. We have studied the output
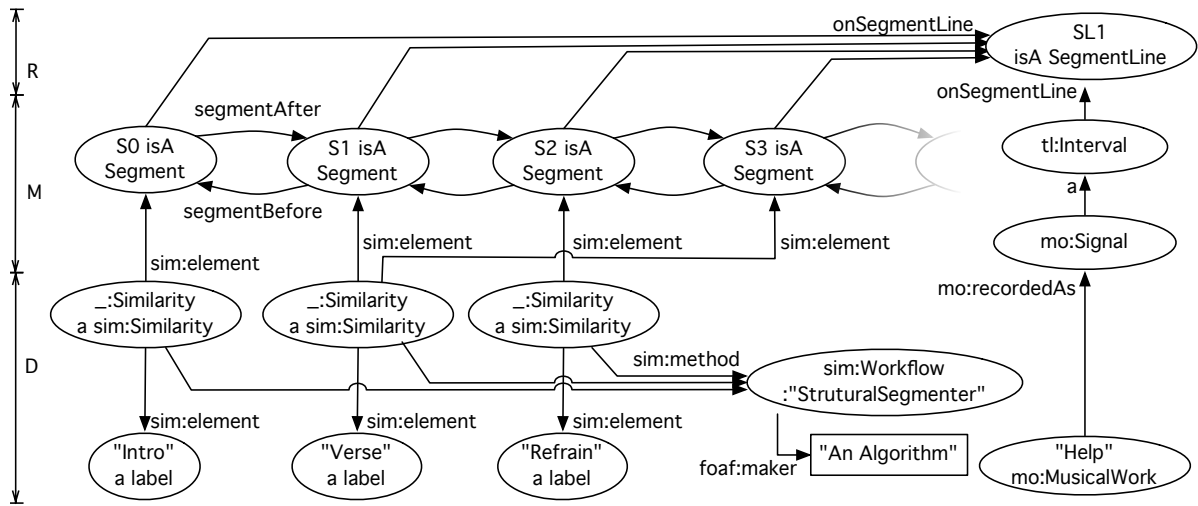
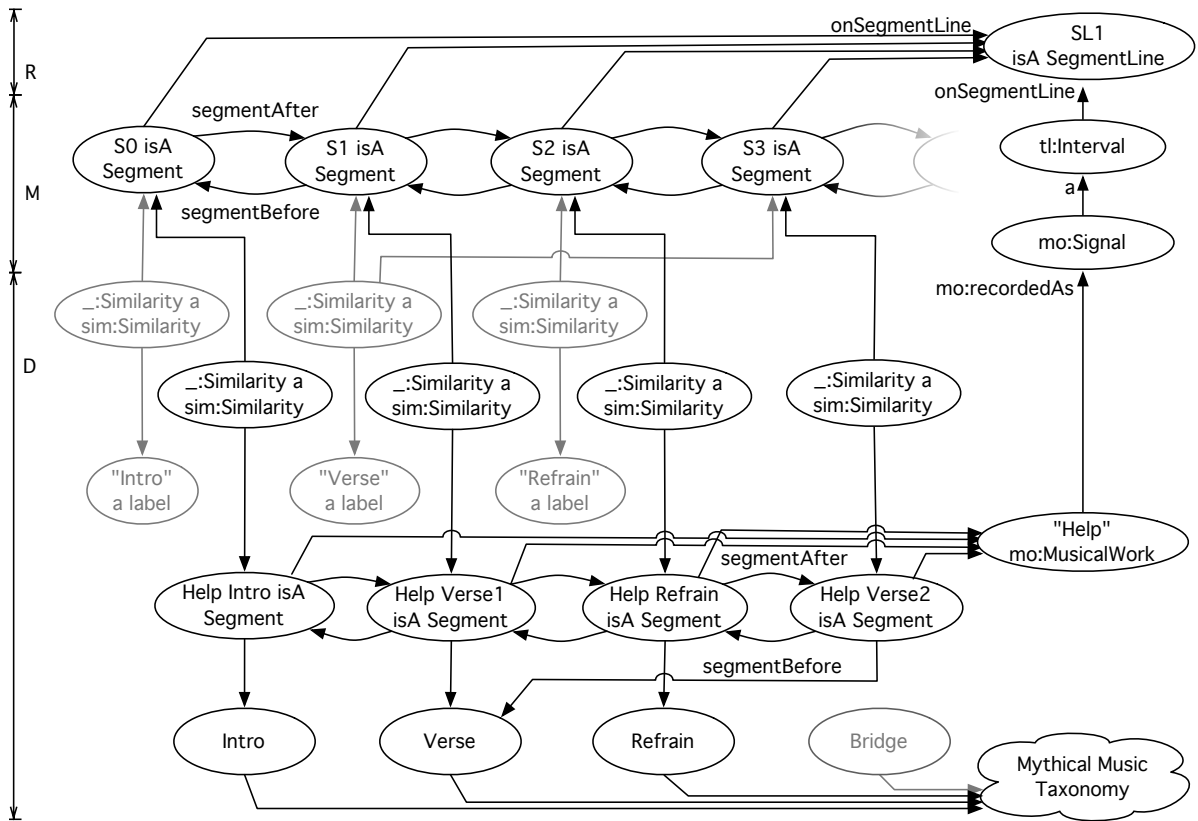**Fig. 4**. Structural segmentation modeled with a relative SegmentLine.

Fig. 5. Extending Figure 4 to express relationships to musicological concepts.

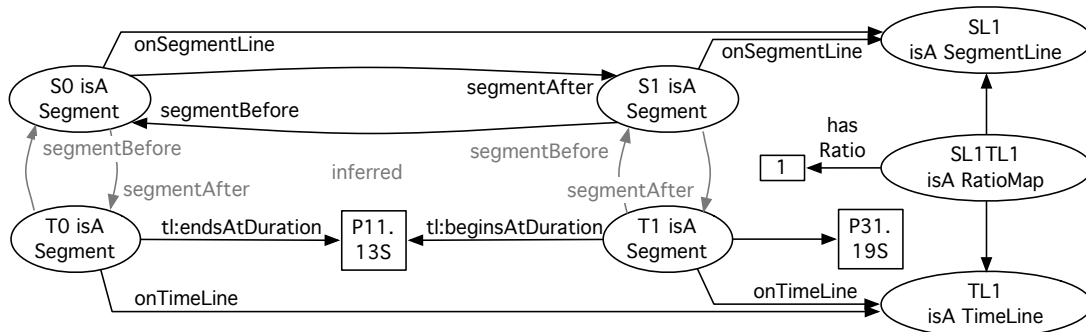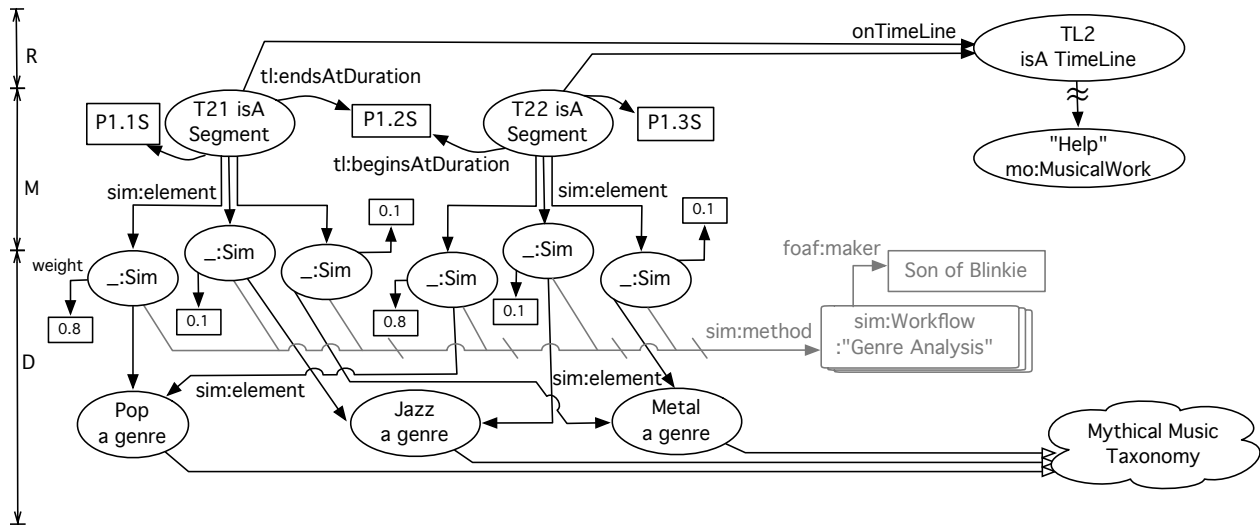**Fig. 6**. Two sets of segments, one on a TimeLine and one a SegmentLine, with RatioMap between them.

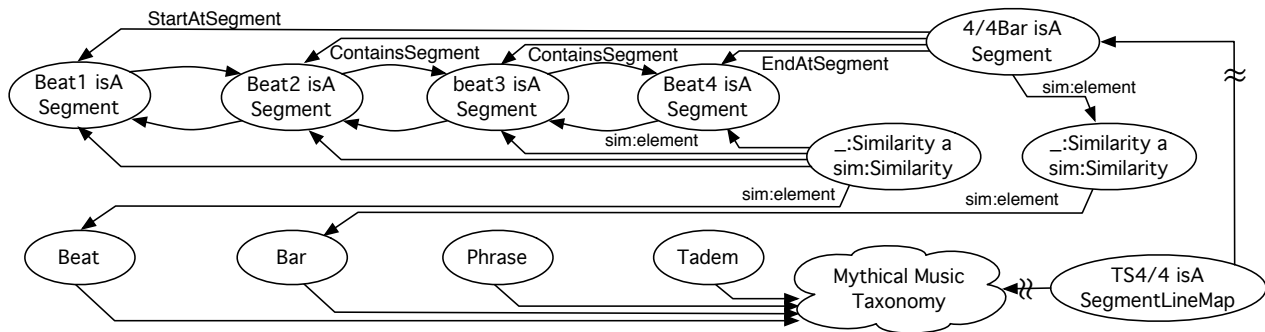**Fig. 7**. Using the Segment Ontology to model frame-by-frame genre analysis.

Fig. 8. Using the Segment Ontology to model the relationship between beats and bars in a $\frac{4}{4}$ time signature.

formats used in MIREX 2010, and in this example look at the structural segmentation task [3]. MIREX algorithmic output for the tasks is stored in the format of the following example snippet (three column, plain text, tab separated, one segment per line):

```
0.000      5.223      A
5.223      15.101     B
15.101     20.334     A
```

With the columns denoting onset time, offset time and label respectively.

To model MIREX output we use the same model shown in Figure 3 with the following annotations in the figure: %A, %B, %C are the "A", "B", "C" from wiki format (note that the "C" is a section not encapsulated in the snippet); "*" indicates a property not explicitly in the MIREX encoding: task (sim:Workflow); algorithm (foaf:maker); digital audio header info such as PCM, 16-be, 44100 Hz, and mono (mo:recodedAs, mo:Signal, tl:Interval); and the musical work itself (mo:musicalWork).Furthermore, the richer data model presented by
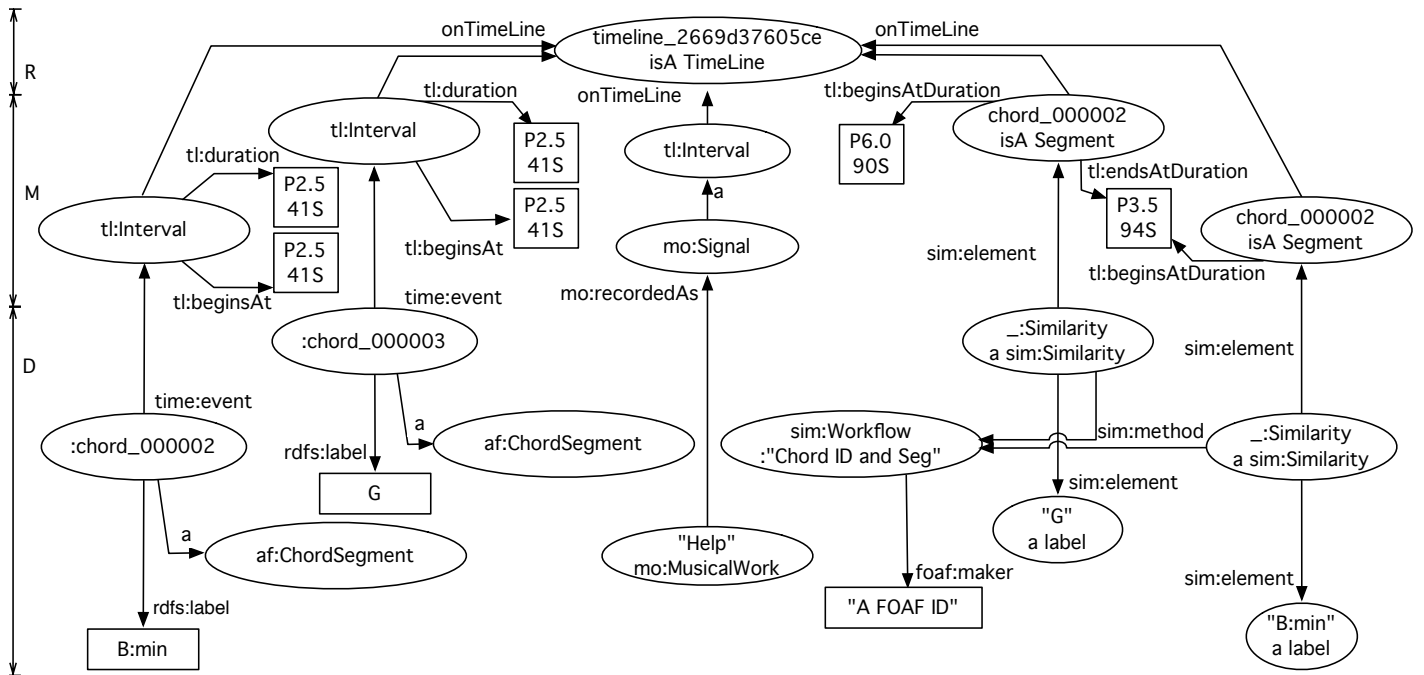
Fig. 9. A side-by-side comparison of the Chord plus AF versus Segment approach. Note both use TL.

our approach allows joining this structural segmentation data with domain data and other analyses performed on the same timeline, enabling secondary information discovery.

6.3. Chord Ontology

The Chord Ontology has its roots as a task-specific ontology (e.g. [9]) and as such captures both domain-specific and music-generic terms; as it builds upon MO and TL it is the domain-specific modeling of chords that is of particular interest – it is an important test that the Segment Ontology and approach can include and support ontologies such as the Chord. In Figure 9 we show a side-by-side comparison of a native encoding (again, from "Help!") using the Chord Ontology, and then the same information successfully represented using the Segment Ontology.

7. CONCLUSIONS AND FUTURE WORK

We have presented the Segment Ontology, that maximally enables post-hoc joins across other datasets and domain structures through a separation of concerns between segmentation, structure, and musical domains. The next stage of our research is to deploy the ontology across the collections of ground truth and computational analysis discussed in section 1, integrated as a part of a Linked Data API through which researchers can access, study, and manipulate results.

Completion of the Mythical Music Taxonomy is left as an exercise for the reader.

8. REFERENCES

[1] David De Roure, J. Stephen Downie, and Ichiro Fujinaga, "Salami: Structural analysis of large amounts of music information," in *UK e-Science All Hands Meeting 2010, Software Sustainbility Workshop*, May 2010.

[2] Kevin R. Page, Ben Fields, Bart J. Nagel, Gianni O'Neill, David C. De Roure, and Tim Crawford, "Semantics for music analysis through linked data: How country is my country?," in *e-Science (e-Science), 2010 IEEE Sixth International Conference on*, 2010, pp. 41 –48.

[3] Christophe Rhodes, Tim Crawford, Michael Casey, and Mark d'Inverno, "Investigating music collections at different scales with AudioDB," *Journal of New Music Research*, vol. 39, no. 4, 2010.

[4] Samer Abdallah, Yves Raimond, and Mark Sandler, "An ontology-based approach to information management for music analysis systems," in *Audio Engineering Society Convention 120*, 5 2006.

[5] Yves Raimond, Samer Abdallah, Mark Sandler, and Frederick Gaisson, "The Music Ontology," in *Int. Conference on Music Information Retrieval*, 2007.

[6] Kurt Jacobson, Yves Raimond, and Mark Sandler, "An ecosystem for transparent music similarity in an open world," in *10th Conference of the International Society of Music Information Retrieval*, 2009.

[7] Chris Cannam, Christian Landone, Mark Sandler, and Juan Pablo Bello, "The sonic visualiser: A visualisation platform for semantic descriptors from musical signals," in *Proc. of Int. Conference on Music Information Retrieval*, 2006.

[8] Gyorgy Fazekas, Yves Raimond, Kurt Jacobson, and Mark Sandler, "An overview of semantic web activities in the OMRAS2 project," *Journal of New Music Research*, vol. 39, no. 4, 2010.

[9] Matthias Mauch and Simon Dixon, "Simultaneous estimation of chords and musical context from audio," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1280–1289, 2010.