Structural Analysis of Large Amounts of Music Information

Mert Bay, John Ashley Burgoyne, Tim Crawford, David De Roure, J. Stephen Downie, Andreas Ehmann, Benjamin Fields, Ichiro Fujinaga, Kevin Page, and Jordan B. L. Smith

1   Introduction

In this progress report, we summarize our accomplishments in the past year on the SALAMI (Structural Analysis of Large Amount of Music Information) project. Our focus has been to develop a state-of-the-art infrastructure for conducting research in music structural analysis. The structure of this report as well as the division of our tasks has been divided naturally into three parts: the McGill group mostly worked on the annotation and the creation of ground truth data (Section 2); the Oxford and Southampton group developed a new model of data representation of sequential and hierarchical divisions (Section 3); and the University of Illinois group is building the computational infrastructure to collect the data, test the algorithms, and to perform the massive calculations (Sections 4–6).

2   Ground truth

2.1   Motivation

Before executing the structural analysis algorithms on the several hundred thousand recordings that were assembled for the SALAMI project, we need to provide evidence that the algorithms will succeed at generating reasonable descriptions of each piece's structure. This demands the creation of a human-annotated ground truth dataset to validate and, where necessary, to train the algorithms.

Creating a ground truth dataset is a complex task that raises several issues, foremost among them: how can we assert that the data collected represents the "truth"? We acknowledge—as must anyone studying musical form—that the form of a piece of music is not an empirically measurable feature, but rather a subjective feature that requires some amount of perception and creative interpretation on the part of the listener. Nevertheless, the study of form attests to the fact that with shared training, different listeners can agree to a considerable extent on how to describe the form of pieces.

This section describes important attributes of the ground truth dataset that was collected, including: the provenance and genre of the pieces included; the annotation format used to encode the descriptions; and the annotation procedure employed. The account below includes some of the main reasons for the design of the database, but a fuller justification and the timeline of the project appear in the following section.

2.2   Description

The choice of recordings to include was influenced by the goals of the project and the practicality of assembling and annotating a large collection of works.

One of SALAMI's major goals was to provide structural analyses for as wide a variety of music as possible. Whereas previous annotated databases of structural descriptions had

generally focused on studio recordings of popular music, with an additional few focusing on classical music, the SALAMI database should also include jazz, folk, the music of cultures from across the globe, known colloquially as "world" music, and live recordings. The ground truth dataset includes a representative sample of music from all these genres. The final composition of the database according to these genres is shown in Table 1.

Table 1 The number of annotated pieces by genre

	Double-keyed	Single-keyed	Total	Percentage
Classical	159	66	225	16%
Jazz	225	12	237	17%
Popular	205	117	322	23%
World	186	31	217	16%
Live recordings	273	109	382	28%
Total	**1048**	**335**	**1383**	100%

Double keying refers to collecting two independent annotations per recording. The majority of pieces are double-keyed, but in some cases single keying was appropriate. Most importantly, roughly 120 of the single-keyed pieces belong to other widely-used databases of structural annotations: the RWC (Goto et al. 2002) and Isophonics[1] collections. Single keying these files allows us to economically compare our results with those of others.

It would be difficult to maintain the correct proportion of genres if recordings were collected from a database such as the Internet Archive[2] with limited and inconsistent metadata. Therefore most of the recordings were collected from Codaich (McKay et al. 2006), a large database with carefully curated metadata, including over 50 subgenre labels which can be categorized under the four domain labels used here. The live recordings are all gleaned from the Internet Archive. While the genre of each of these recordings is not known, the majority appears to be in the popular and jazz categories.

The project hired nine annotators who contributed on average 270 annotations. Each annotator had a B.A. in music and was pursuing either a M.A. or Ph.D. in Music Theory, or a Ph.D. in Composition at McGill University.

2.3 Annotation format

Musicological or music theoretical analyses of structure may take many forms, but when algorithms are involved, the possibilities for an annotation format are constrained: for example, while each annotation could consist of a paragraph-length description of the form, this would be of little use to most imaginable algorithms.

---

[1] http://www.isophonics.org/datasets
[2] http://www.archive.org/details/audio

In order that the annotation format be machine readable, we limited the type of information that the descriptions will contain, yet it was designed to be able to describe the form of virtually any kind of music. Because the annotations were created by humans, the annotations were also designed to be easily writeable and readable by humans.

The most important information in our annotations is the segmentation of the recording into sections, and the segment labels that indicate which are similar to or repetitions of one another.

Most structural annotations encode this information in a very simple format: each segment boundary time is enumerated along with its label. As pointed out by Peeters and Deruty (2009), however, these labels may be inconsistently applied due to the conflation of the musical surface, the function of a particular passage, and the instrumentation. For instance, an introduction section that is repeated as a closing may receive two distinct labels: "intro" and "outro." Previous corpora of structural annotations that suffer from this ambiguity, such as the Center for Digital Music's Beatles annotations,[3] may not be helpful for validation purposes. However, other corpora such as RWC (Goto et al. 2002) use a vocabulary that is too highly constrained to be applicable to all the genera of music included in SALAMI.

Peeters and Deruty proposed a novel annotation format, which uses a set vocabulary of 21 labels that distinguish between the "musical similarity" between sections, and the "musical role" and "instrument role" of each section. We adopted this tripartite distinction, but over the course of testing made several modifications to suit our purposes.

The final annotation scheme consisted of separate tracks for *musical similarity, function,* and *lead instrument:*

The *musical similarity* track consisted of two annotations at different scales (large and small), one finer-grained than the other, each identifying which portions of the recording use similar musical ideas. Simple letter labels were used; the large-scale track generally used five or fewer labels, while the small-scale track could use as many labels as necessary. Special labels indicate silence ("s") and non-music, such as applause or banter in a live recording ("Z"). Varying degrees of similarity could be identified using prime symbols ('). Every portion of the recordings was labeled in both large- and small-scale tracks.

A separate *function* track, generally aligned with the large-scale segment boundaries, provides function labels where appropriate. The possible labels were drawn from a strictly limited vocabulary of roughly 20 labels. Some of these labels express similar functions and can be grouped together if desired: for example, "pre-verse," "pre-chorus," "interlude," and "transition" all express similar functions and could, if desired, all be re-labeled as "transition."

A separate *lead instrument* track, generally aligned with the small-scale segment boundaries, indicates wherever a single instrument or voice takes on a leading, usually melodic role. The vocabulary for these labels was not constrained, and unlike the other tracks, lead instrument labels could potentially overlap, as in a duet. Note that as with the function track, there may be portions of the recording with no lead instrument label.

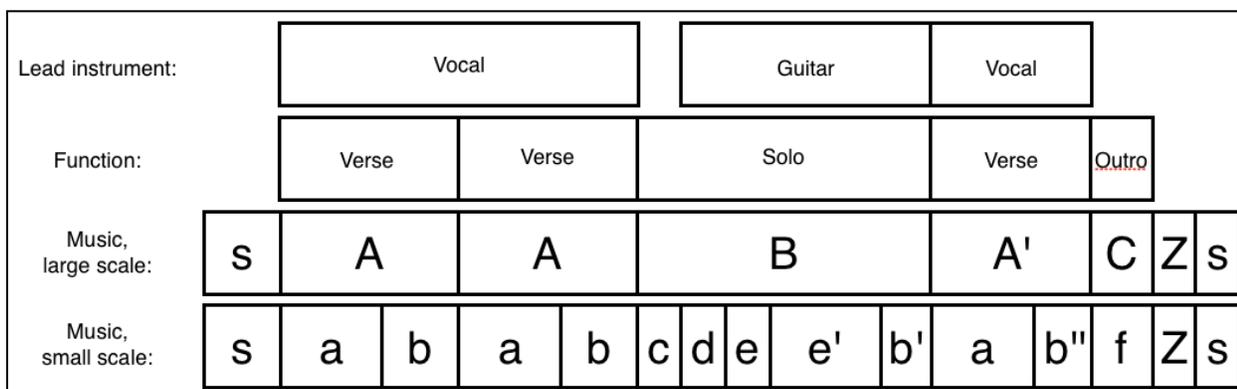A graphical example of the annotation scheme is shown below.

---

[3] http://www.isophonics.org/content/reference-annotations

Figure 1: An example of musical structure of a piece

In the written format devised for this scheme, the example in Figure 1 would begin as:

```
0.000     silence
8.145     verse, A, a, (vocal
20.31     b
29.04     verse, A, a
41.74     b, vocal)
49.82     B, c, solo
56.20     d
etc.
```

2.3.1   Function label vocabulary

The following function labels are permitted: *introduction, verse, chorus, bridge, instrumental, solo, transition, interlude, pre-chorus, pre-verse, head, main theme, (secondary) theme, exposition, development, recapitulation, outro, coda, fadeout, silence,* and *end.* Working definitions for each term are specified in our Annotator's Guide (see Figure 2 for a summary).

Note that some of the labels are genre-specific alternatives to others: for example, the *head* in jazz song is analogous to a *chorus* in a pop song or a *main theme* in some classical genres. Additionally, some subsets of the vocabulary can function as synonym-groups that can be collapsed onto a single function label if desired. For example, while our Annotator's Guide suggests a fine distinction between *pre-chorus, pre-verse, interlude,* and *transition* sections, they are all synonyms of *transition.* Specifying these groups enables someone wanting to train an algorithm on the SALAMI data to observe these distinctions or collapse the synonym group onto a single label.

Together, the terms *exposition, development,* and *recapitulation* are specific to sonata form and may in special cases be used to annotate a third level of structural relationships on a scale larger than the usual large-scale labels. However, *development* also has wider applicability and may be used to label the function of a "contrasting middle" section in many contexts, from various classical genera to progressive rock.

The vocabulary is separated into various categories below. The ***instrumental, transition*** and ***ending*** groups are all synonym groups. The ***genre-specific alternatives*** are analogous to the ***basic functions*** but are not specific to popular music. The ***form-specific alternatives*** are

especially included for certain classical forms, although among these the term *development* has broader use. Note that in the **ending** group, the label *fadeout* is a special label that can occur in addition to any other label. For example, if the piece fades out over a repetition of the chorus, then the last section may be given both labels: *chorus* and *fadeout.*

Figure 2: Summary of label vocabulary

2.4 Annotation workflow

The annotation format and data collection took place over the course of 10 months, although most of the data was collected within the first 16 weeks.

First, previous annotation formats and databases of annotations were researched. Potential annotation formats were devised and tested by the project leaders, and a tentative format was set at the end of two months.

Next, candidate annotators were trained in the annotation format and in the Sonic Visualiser environment (Cannam et al. 2006), which was used to make the annotations. Candidates who were able and willing to continue with the project were hired and data collection began the following week.

Because the annotation format had not been tested on a significant scale before work began in earnest, the first six weeks of data collection were conceived as an extended trial period. Every week or two, annotators were given a new batch of assignments in a new genre, beginning with popular, which was expected to be the least problematic, and continuing in order with jazz, classical, and world, which were predicted to be of increasing difficulty.

After each new assignment, we solicited feedback from the annotators on difficult pieces they encountered and weaknesses or ambiguities in the annotation format that were revealed. Group meetings were held so that these general problems could be discussed. Based on the feedback, some annotation rules were changed (e.g., the function label vocabulary expanded or contracted), and new heuristics were introduced (e.g., we introduced a preference to have segment boundaries fall on downbeats even in the presence of pickups). In at least one case, a

major revision of the format originated from annotator feedback: our original annotation format used a single music similarity track with some hierarchical information embedded, but early on we switched to the two-track system described in the previous section.

At the end of the six weeks, supervision of the annotators was relaxed and any problems addressed on an *ad hoc* basis. Data collection continued over the next 12 weeks, by which point the majority of assignments had been completed. The median transcribing time was 15 minutes per track and majority of transcribing took between 10 and 25 minutes. In general, more time was needed for Classical and World music than Popular and Jazz music but this may be attributed to the generally longer time of the former group.

3 Data representation: Segment Ontology

3.1 Introduction and background

Existing semantic representations of music analysis encapsulate narrow sub-domain concepts and are frequently scoped by the context of a particular Music Information Retrieval (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.

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 (Page et al. 2010). In the context of SALAMI project and in an effort to model the segmentation task itself in more detail, and to 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 section 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, and then present some examples.

3.2 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 Rhodes et al. (2010).

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.

While supporting other domain-specific categorizations is a motivating use-case for the segment ontology, we explore the two most directly applicable to existing MIR systems: musicology and MIR tasks. 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, and bridge. 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.

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.

3.3 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 (Abdallah et al. 2006). 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 our 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 (Raimond et al. 2007). 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) was conceived to model music similarity (Jacobson et al. 2009). The current version's use of blank 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.

3.4   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 3 and detailed below:

Segment: an Interval with the addition of a "label" expressing an association (SIM) that can be "placed" upon TimeLines (TL) and SegmentLines. There are five intraSegment 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: an AbstractTimeLine and a relative complement to the temporal 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 NonSequential-Map, the least specified, whereby sequential order of Segments is not preserved across SegmentLines.

Thus, the segment ontology encodes the high-level relationships and music-generic concepts introduced in Section 3.2. Domain-specific annotations, such as MIR-task and musicology, will be described independently using appropriate ontologies. We model the relationships that stem from these as domain-specific terms in the same way: as (associative) annotations to Segments, SegmentLine, and TimeLines, and the high-level relationships between them.

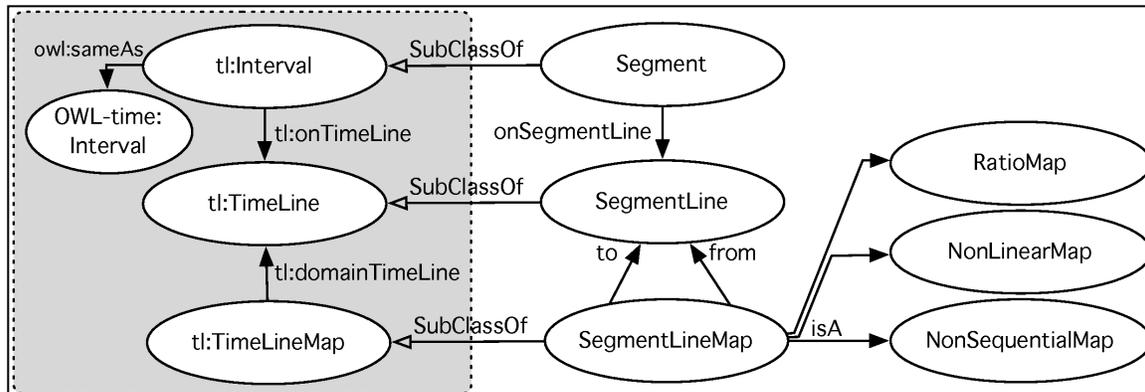**Figure 3: The class structure of Segment Ontology. Concepts from TimeLine ontology are on the grey background.**

3.5   Examples

Throughout these examples we reference and compare an existing analysis of the Beatles "Help!" Figure 4 is a generic visualization of the analytic structures that can be found in this piece of music; it is worth recognizing that although Figure 4 does not use any specific ontology or data structure, it does invoke a temporal dimension most would apply as their default interpretation.
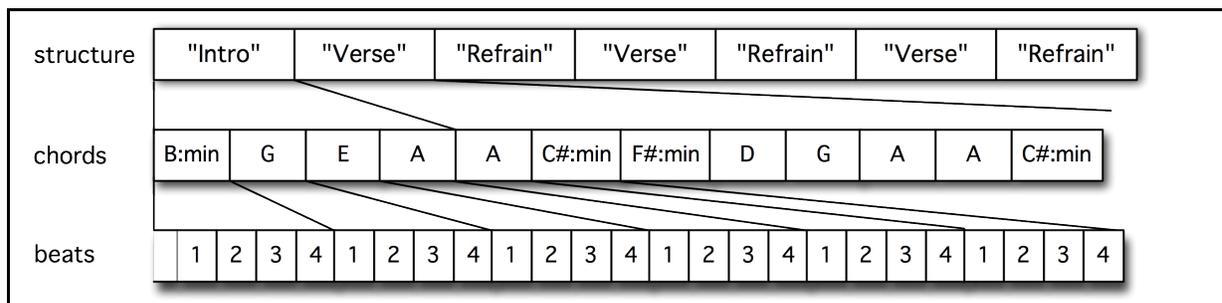

Figure 4: Segmentation of the song "Help!" by The Beatles by song structure, chord, and beat, with alignment shown.

In these examples we have also arranged the models according to the categorization introduced in Section 3.2 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.
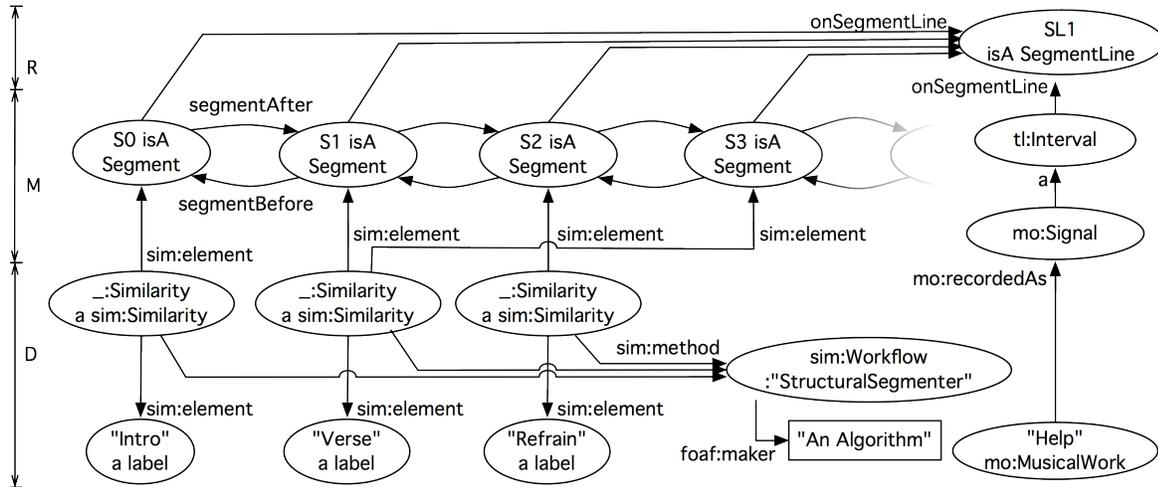
Figure 5: Structural segmentation modeled with a discrete TimeLine

Figure 5 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 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 6 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 5. Note that the segments are just given a label (e.g., "Verse" or "Refrain" but with no meaning).
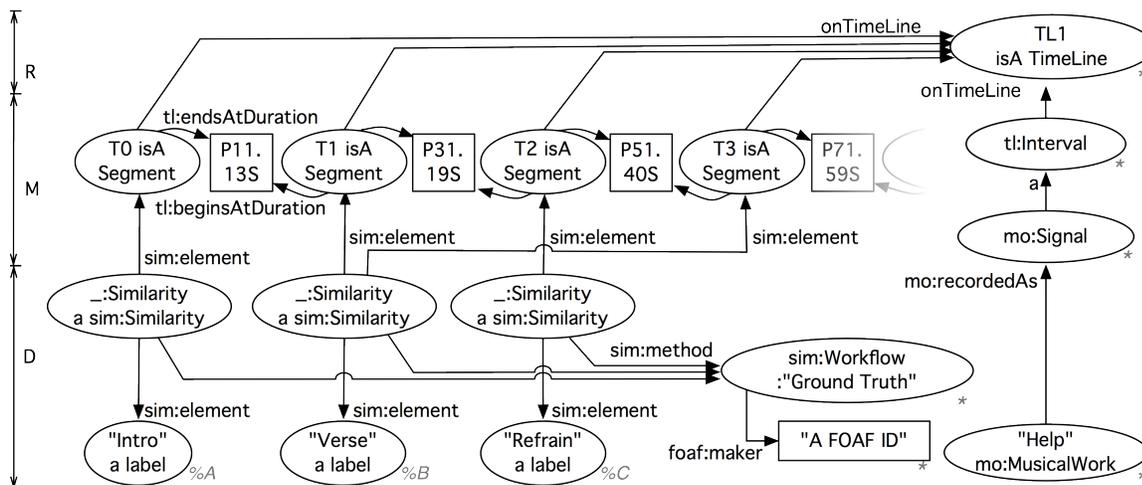

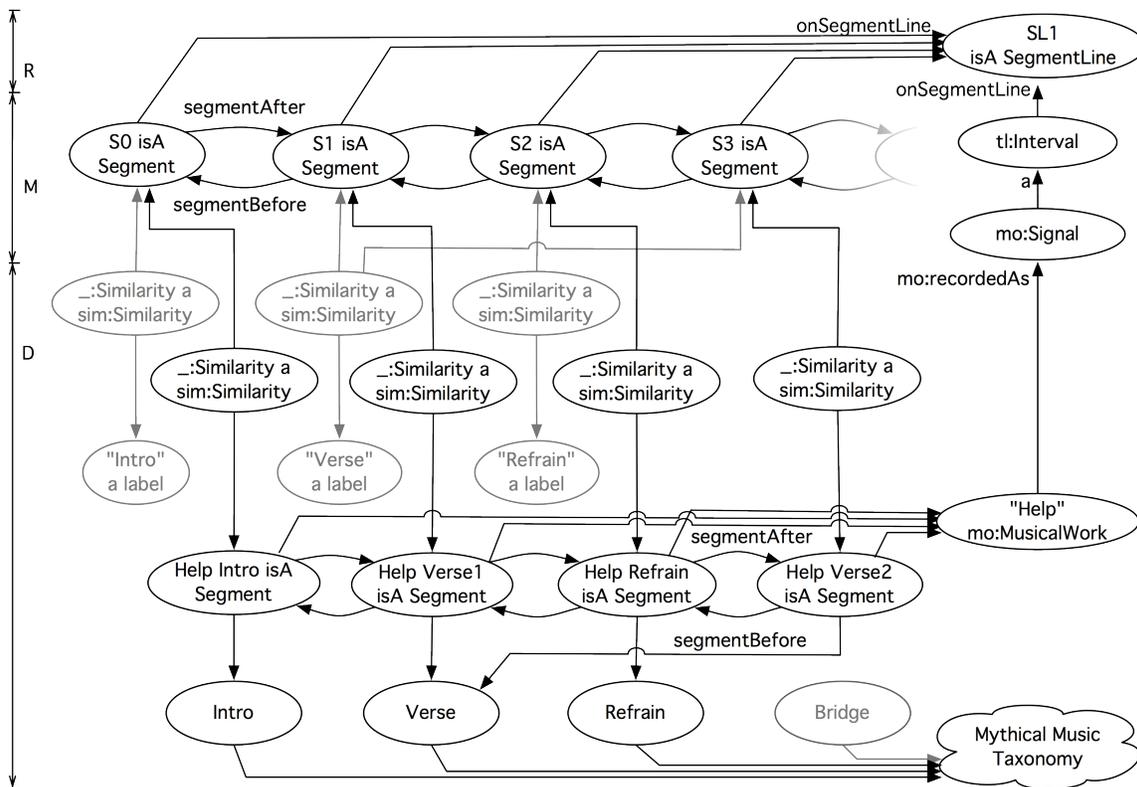**Figure 6: Structural segmentation modeled with a relative SegmentLine**

Figure 7: Extending Figure 6 to express relationships to musicological concepts

Figure 7 relates segmented analysis to musical concepts, an extension of Figure 6 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).

4   Algorithm evaluation

As part of an additional effort to supplement the evaluation results over the course of the last two MIREX structure evaluations,[4] five structural analysis algorithms were run and evaluated against a set of over 1,000 songs annotated at McGill. The average processing time for each of the algorithms is shown in Table 2. To evaluate the algorithms, a broad range of metrics exist. For brevity, we present only frame-pair clustering (FPC) (Levy and Sandler, 2008). Both the algorithm result and ground truth are divided into short time frames (e.g. 100 ms). All pairs of frames are subsequently analyzed. The pairs in which both frames share the same label (i.e. belong to the same cluster) form the sets P_E (for the system results) and P_A (for the ground truth). We can therefore calculate the pair-wise precision, P; recall, R; and F-Measure, F, as follows:

---

[4] http://www.music-ir.org/mirex

$$P = \frac{|P_E \cap P_A|}{|P_E|} \qquad R = \frac{|P_E \cap P_A|}{|P_A|} \qquad F = \frac{2PR}{P + R}$$

The overall evaluation results show a correspondence with previous MIREX evaluations. Most algorithms tend to annotate at a coarser level of hierarchy. Moreover, since each musical piece has multiple annotations, we are able to evaluate how closely two humans come to agreement on the structural annotation of a piece. The evaluation results of a selection of three algorithms and the human-to-human evaluation can be seen in Table 3.

The evaluations tend to enforce two findings: First, human-to-human agreement is still higher than algorithm-to-human agreement (Frame pair clustering F-measure of 0.721 vs. 0.565, respectively). This leads to the conclusion that structural annotation by machines is still not a solved problem. Secondly, although humans "outperform" machines currently, human-to-human evaluations also indicate that there is quite a bit of disagreement between human expert annotators on how pieces should be structurally segmented. It is this finding that reinforces our belief for the SALAMI project that musical pieces should be annotated by as many "experts" as possible (in this case machine experts). It is from the opinions of multiple sources that we believe the most benefit can be drawn.

Table 2: Structural analysis processing time by different algorithms

Algorithms	Average processing time (min. / piece)
WB1 (Weiss and Bello 2010)	2.28
GP7 (Peeters 2007)	2.64
BV1 & BV2 (Sargent et al. 2010)	2.94
MND1 (Mauch et al. 2009)	5.60
MHRAF2 (Martin et al. 2009)	6.38

Table 3: Evaluations of three algorithms and a human against a ground truth

Algorithm	FPC F-measure	FPC Precision	FPC Recall
Human	0.7211	0.7692	0.7453
MHRAF2	0.5647	0.6319	0.5782
MND1	0.5590	0.6611	0.5848
WB1	0.5522	0.5928	0.6091

5 Interactive visualizer

An important aspect of the SALAMI project is to allow users and the community to explore and interact with the structural annotations generated for a large music digital library. To this end, an interactive visualizer with structurally aware music playback has been developed. The visualizer, seen in Figure 8, plots all available annotations for a given musical piece. The plot represents a timeline of the piece with each labeled rectangular segment corresponding to a structural segment of the piece. The visualization can be zoomed and panned. Moreover, clicking on a segment will play the portion of the audio corresponding to that segment. Therefore users can quickly browse similarly labeled segments to find important repetitions, themes, etc.
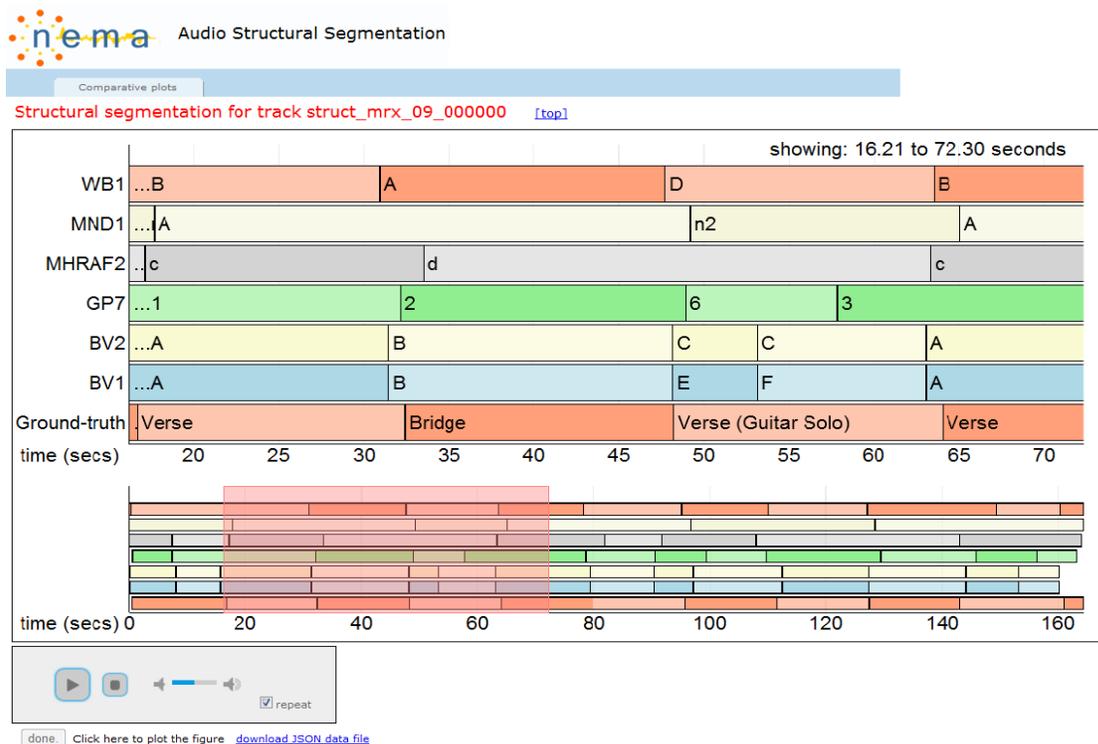

Figure 8: A screenshot of SALAMI's interactive visualizer and audio player interface for exploring multiple structural analyses

6 Work in progress

The SALAMI project is currently in the process of executing its main goal, namely the annotation of hundreds of thousands of music pieces by multiple machine experts. This goal represents a significant resource management problem. Each algorithm, on average, spends one to five minutes of compute time to annotate a single piece of music. Therefore the annotation of, for example, 200,000 pieces by five different algorithms requires roughly five to six years of compute time. Leveraging available supercomputing infrastructure is the only means to achieve this computational goal in a short amount of time.

However, modern supercomputing infrastructures pose some additional problems over evaluating current structure algorithms on smaller datasets, as has been done to date. Firstly,

13

most structure algorithms are in the research development stage and are not commercial-grade code. With little access and availability to install custom libraries on a supercomputing cluster, each existing algorithm must be packaged such that it is a completely independent and platform-agnostic entity. Secondly, audio data, even compressed, represents a fairly large disk storage challenge. Persistent data storage of large amounts of data is not available on most shared supercomputers.

To address these challenges, each structure algorithm has now been bundled with all necessary libraries and dependencies and scripted such that it represents a platform-independent object with no need for external libraries or compute engines (e.g., MATLAB) to be installed on the cluster. Additionally, the entire SALAMI collection has been migrated to a persistent tape mass storage device. The audio data is in a lossy compressed format and its current total is 500 GB (200,000 tracks). The audio data will be fetched as needed during computation, decompressed on the cluster end, and the algorithms run against the decompressed raw audio. Decompressing on the supercomputing side means data can be transferred more quickly at the expense of some computation time in uncompressing the data. The SALAMI team is currently negotiating the actual supercomputer that will be used for the runs (possibilities are at Illinois, Tennessee, and San Diego).

7 Conclusions

As one of the first experiments in large-scale music data mining, we have made tremendous progress by creating a large amount of high-quality annotation data and in modeling the data structure needed for this type of time-based hierarchically organized data stream, in our case, music. Furthermore, based on our experiences in running the annual MIREX evaluations, we were able to relatively quickly construct a robust infrastructure to run the large-scale experiment. In the forthcoming months we will execute the "Big Run," which involves running several structural analysis algorithms on over 200,000 music pieces. Through this work we are establishing a methodology for MIR at large scale, and establishing practices which we hope will enable this research to be continued beyond the immediate lifetime of the project.

8 References

Abdallah, S., Y. Raimond, and M. Sandler. 2006. An ontology-based approach to information management for music analysis systems. In *Audio Engineering Society Convention* 120: 5.

Cannam, C., C. Landone, M. Sandler, and J. P. Bello. 2006. The Sonic Visualiser: A visualisation platform for semantic descriptors from musical signals. In *Proceedings of the International Conference on Music Information Retrieval,* 324–7.

Goto, M., H. Hashiguchi, T. Nishimura, and R. Oka. 2002. RWC Music Database: Popular, classical, and jazz music databases. In *Proceedings of the International Conference on Music Information Retrieval*, 287–8.

Jacobson, K., Y. Raimond, and M. Sandler. 2009. An ecosystem for transparent music similarity in an open world. In *International Society of Music Information Retrieval Conference*.

Levy, C., and M. Sandler. 2008. Structural Segmentation of Musical Audio by Constrained Clustering. *IEEE Transaction on Audio, Speech, and Language Processing* 16 (2): 318–26.

Martin, B., M. Robine, and P. Hanna. 2009. Musical structure retrieval by aligning self-similarity matrices. In *Proceedings of the International Society for Music Information Retrieval Conference,* 483–8.

Mauch, M., K. C. Noland, and S. Dixon. 2009. Using musical structure to enhance automatic chord transcription. In *Proceedings of the International Society for Music Information Retrieval Conference,* 231–6.

McKay, C., D. McEnnis, and I. Fujinaga. 2006. A large publicly accessible prototype audio database for music research. In *Proceedings of the International Conference on Music Information Retrieval,* 160–3.

Page, K. R., B. Fields, B. J. Nagel, G. O'Neill, D. C. De Roure, and T. Crawford. 2010. Semantics for music analysis through linked data: How country is my country? In *IEEE Sixth International Conference on e-Science,* 41–8.

Peeters, G. 2007. Sequence representation of music structure using higher-order similarity matrix and maximum likelihood approach. In *Proceeding of the International Conference on Music Information Retrieval.*

Peeters, G., and E. Deruty. 2009. Is music structure annotation multi-dimensional? A proposal for robust local music annotation. In *Proceedings of the International Workshop on Learning the Semantics of Audio Signals*, 75–90.

Raimond, Y., S.Abdallah, M. Sandler, and F. Gaisson. 2007. The Music Ontology. In *International Conference on Music Information Retrieval.*

Rhodes, C., T. Crawford, M. Casey, and M. d'Inverno. 2010. Investigating music collections at different scales with AudioDB. *Journal of New Music Research* 39 (4): 337–48.

Sargent, G., F. Bimbot, and E. Vincent. 2010. Un système de détection de rupture de timbre pour la description de la structure des morceaux de musique. In *Proceedings of Journées d'Informatique Musicale,* 177–86,

Weiss, R. J., and J. P. Bello. 2010. Identifying repeated patterns in music using sparse convolutive non-negative matrix factorization. In *Proceedings of the International Society for Music Information Retrieval Conference.*