Towards Web-Scale Computational Musicology: An Update on the SALAMI Project

Mert Bay[1], John Ashley Burgoyne[2], Tim Crawford[3], David De Roure[4], J. Stephen Downie[1], Andreas Ehmann[1], Benjamin Fields[3], Ichiro Fujinaga[2], Kevin Page[4] and Jordan B. L. Smith[2]

[1]Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign, US
[2]Schulich School of Music, McGill University, Canada
[3]Department of Computing, Goldsmiths University of London, UK
[4]Oxford e-Research Centre, University of Oxford, UK

1 Introduction

We summarise our accomplishments on the *Structural Analysis of Large Amount of Music Information* (SALAMI) project in which we have developed a state-of-the-art infrastructure for conducting research in music structural analysis. The project was outlined in AHM 2010 with a proof of concept case study, and in the subsequent 12 months we have substantially developed three areas which we outline in the following sections: annotation and the creation of ground truth data, development of a new model of data representation of sequential and hierarchical divisions, and testing the algorithms which will be used to perform the calculations. The ground truth data and the ontology are contributions to the field which have broad applicability beyond this project. In the next phase of the project we will use the algorithms to perform the large-scale analysis.

2 Ground truth

The project demands the creation of a human-annotated ground truth dataset to validate and, where necessary, to train the structural analysis algorithms that will be applied to the several hundred thousand recordings that were assembled for the SALAMI project. 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 also includes 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%

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. We adopted the scheme introduced by Peeters and Deruty [1], making several modifications to suit our purposes. A graphical example of the annotation scheme is shown below.

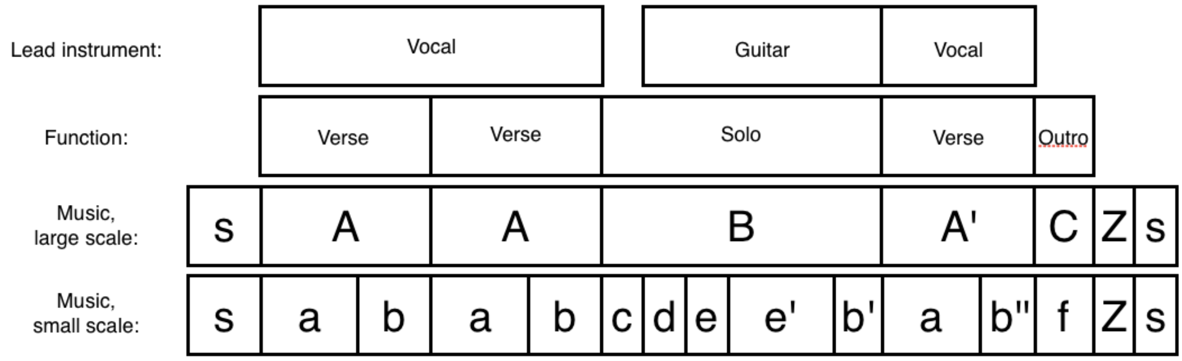

Figure 1: An example of musical structure of a piece

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.
```

3 Segmentation Ontology

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 Music Information Research (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]. 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 have developed the Segmentation Ontology, focused on modeling division of temporal-signal (principally music) into subunits [3].

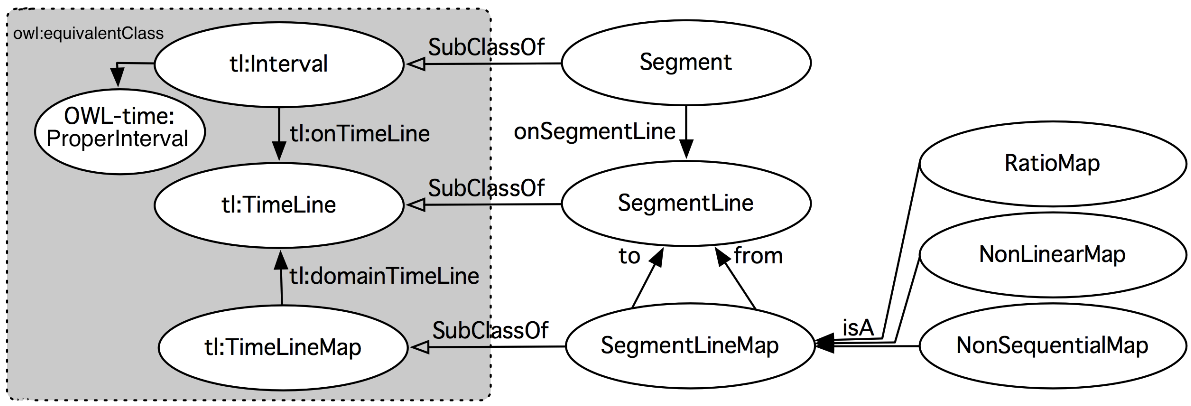

Figure 2: Class structure of Segmentation Ontology

The class structure is shown in Figure 2. 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. Our 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: 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.

4. Algorithm evaluation

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.

Five structural analysis algorithms were run and evaluated against a set of over the annotated dataset. 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 2. 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.

Table 2: Evaluations of 3 algorithms and 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

References

[1] 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.

[2] 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.

[3] Ben Fields, Kevin Page, David De Roure and Tim Crawford (2011) "The Segment Ontology: Bridging Music-Generic and Domain-Specific" in 3rd International Workshop on Advances in Music Information Research (AdMIRe 2011) held in conjunction with IEEE International Conference on Multimedia and Expo (ICME), Barcelona, July 2011.