

# Facilitating Music Information Research with Shared Open Vocabularies

Alo Allik, György Fazekas, Simon Dixon, and Mark Sandler

Queen Mary University of London

{alo.allik,gyorgy.fazekas,simon.dixon,mark.sandler}@eecs.qmul.ac.uk

**Abstract.** There is currently no agreement on common shared representations of audio features in the field of music information retrieval. The Audio Feature Ontology has been developed as part of a harmonised library of modular ontologies to solve the problem of interoperability between music related data sources. We demonstrate a software framework which combines this ontology and related Semantic Web technologies with data extraction and analysis software, in order to enhance audio feature extraction workflows.

**Keywords:** semantic audio analysis, music information retrieval, linked open data, Semantic Web technologies.

## 1 Introduction

Researchers in audio and music information retrieval increasingly use a common set of feature extraction techniques to characterise audio material, and large data sets of features are released for commercial and scientific use. The development of data sets and research tools however is not governed by shared open vocabularies and common data structures. This raises several issues including the lack of interoperability between music analysis tools and the need for adapting program code and software environments for a variety of different formats. There is growing awareness in the music informatics community of the necessity of common representations using Semantic Web technologies and linked data [5].

We demonstrate a software framework which combines Semantic Web technologies with data extraction and analysis software in order to enhance audio feature extraction workflows and improve interoperability and sustainability. The main interface to the framework is the Sonic Annotator Web Application (SAWA), an online semantic audio analysis interface, which is accessible at <http://www.isophonics.net/sawa/>. The main contribution of this work is the development of the new Audio Feature Ontology. The previous version of the ontology [1] had relatively poor adoption in research communities, due to its limited domain coverage and incomplete tool support.

## 2 Components

Our framework utilises a number of software components which were originally developed during the OMRAS2 project[1]. The software components include a

set of extensible ontologies, a plugin interface for audio feature extraction and annotation, and open source research tools. The Audio Feature (AF) Ontology was created within the framework of the Music Ontology[2], the core of a harmonised library of music related ontologies. AF provides a model for describing acoustical and musicological data, and allows for publishing content-derived information about audio recordings. Its aim is to provide a framework for communication. This ontology is currently being updated in the course of an ongoing research project to provide a better coverage of features commonly used by researchers, to enable better harmonisation with existing software tools, and to support a wider set of data sets and use cases.

Vamp is a plugin system designed for audio feature extraction. It is a modular and extensible collection of shared libraries that are meant to be embedded in a host application. Vamp plugins accept audio data as input and produce structured data as output. The plugin system is supported by the Vamp Plugin Ontology within the context of Semantic Web data. It allows the association of richer metadata with each plugin - for example, information about plugin creators, licensing, or high level content description - and enables the description of plugin structure and configuration. The ontology also provides associations of plugin outputs with terms in the Audio Feature Ontology to express what the output describes.

Sonic Visualiser[3] is a desktop application for visualising audio and a wide range of automatic annotations, including anything from spectrograms to high-level information about musical structure. Sonic Visualiser can handle a wide range of audio formats, and supports the Vamp plugin interface. Sonic Annotator is an audio analysis application which applies Vamp feature extraction plugins to audio data in a batch. It is built using Sonic Visualiser libraries. The two applications therefore share capabilities such as broad support for different audio file formats, network retrieval of audio files, and the interpretation of feature extraction specifications in RDF using the Vamp Plugin Ontology.

Sonic Annotator Web Application (SAWA)[4] is a framework that ties the above described components into a coherent Web-application framework. It utilises Semantic Web vocabularies and ontologies, including the Audio Feature Ontology, which provides its internal model. The main objective of the SAWA framework is to facilitate the use of semantic audio analysis on the Web and support a wide range of different algorithms, while using a flexible, ontology-based common data model for representing information about audio analysis algorithms, configurations and results. SAWA also stores feature extraction results and avoids repeated computation of features given the same audio content and algorithm with the same parameter configuration. SAWA uses the Vamp plugin system to support a wide range of different audio feature extraction algorithms.

### 3 Web-Based Semantic Audio Analysis

This present work focusses on two novel aspects of interoperability within the described collection of components illustrating the practical applications of the

Audio Feature Ontology. First, the SAWA framework is utilised as a Web-based audio analysis system. The SAWA Feature Extractor is a comprehensive and user friendly feature extraction system, allowing the user to compile collections of audio files for batch analysis and to select desired feature plugins, eventually making the results available in multiple formats for download. The second part of the demonstration highlights how the SAWA SPARQL endpoint can be consumed from the Sonic Visualiser desktop application, and how SAWA can be utilised as an optimised data server in music information retrieval workflows.

SAWA is designed with the intention of providing an easy to use human interface as well as a SPARQL query interface. SAWA Feature Extractor works with a collection of audio files uploaded by the user. Optionally, each audio file is identified using a fingerprint, and basic bibliographic metadata given an associated identifier (PUID) is retrieved from the MusicBrainz service<sup>1</sup>. The fingerprint serves as the identifier that is used for storing and later identifying provenance information and to facilitate RDF caching. SAWA Feature Extractor allows for

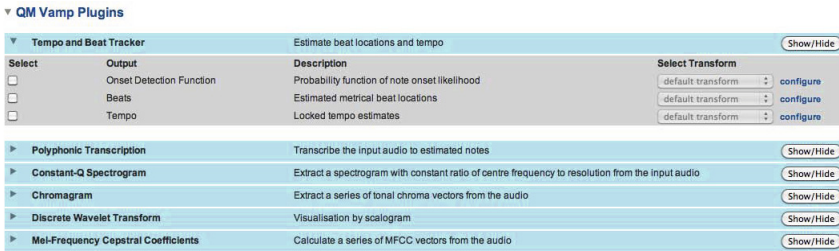


Fig. 1. SAWA interface for Vamp plugins

the selection and configuration of one or more Vamp outputs and the execution of transforms on previously uploaded files. The range of existing Vamp plugins include estimators for musically meaningful features (note onset detectors, beat and tempo estimators, structural segmentation, and key estimators), low-level audio feature extractors, metadata annotators and calculators for dense features that are often used in visualisations, such as chromagram and harmonic spectrum. Results, returned as RDF data, can be examined using an RDF browser, imported in Sonic Visualiser and viewed in context of the audio or published on the Semantic Web.

Figure 1 shows how a transform can be selected from a list of Vamp plugins and how for each plugin a specific set of parameters can be configured for execution. The user may click on the triangular symbols associated with Vamp plugins to expand the content such as the description of each output of a plugin. These transforms are saved temporarily and applied individually on each audio file after submitting a query. Once a set of feature extractors has been selected and configured, a query can be submitted to perform the analysis of all or a user defined

<sup>1</sup> <http://musicbrainz.org/>

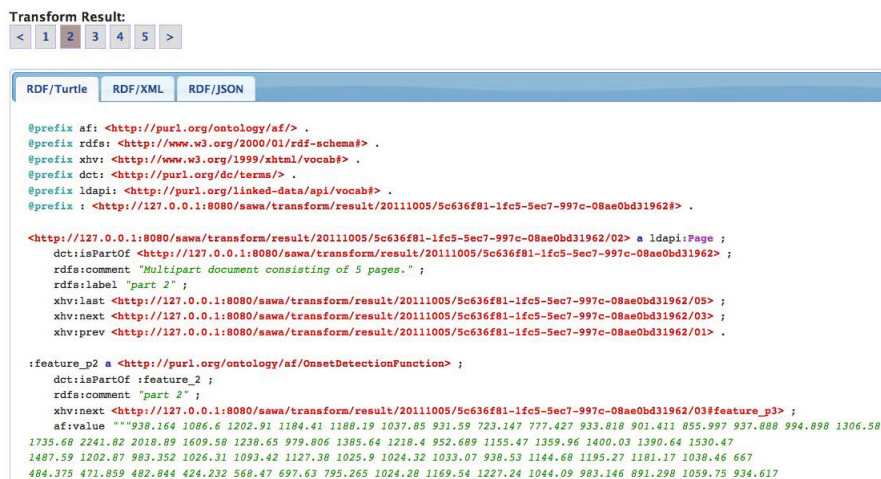


Fig. 2. Paginated feature extraction results in RDF/Turtle format

subset of the uploaded audio files. The progress of the analysis can be monitored using the status display. While the feature extraction process is running, the results are loaded by the Web application as they become available and a table of results is dynamically expanded. The feature extraction results are available in several different formats generated on demand from the default RDF/Turtle syntax. These results can be examined within the application, downloaded as uncompressed RDF or as a ZIP compressed file. Additionally, each result receives a permanent URI. This allows all transform results to be accessed later, even after the user session has expired. The URI is designed to include the date of feature extraction and a universally unique identifier (UUID) generated given the audio fingerprint and the transform configuration in order to avoid potential URI collisions. These permanent links are available in RDF/Turtle and RDF/XML formats. Results consisting of a large number of RDF triples and those that include large matrices are paginated.

The SPARQL-endpoint of SAWA Feature Extractor can be consumed from Sonic Visualiser. The SPARQL client accesses a user defined SPARQL-endpoint and converts the returned data into a format a Vamp host like Sonic Visualiser can understand. The simple query interface of this client is shown in Figure 3. This enables the user to incorporate the functionality of SAWA into the desktop feature extraction workflow and take advantage of the interoperability and efficiency the framework offers. The query results can be linked with other Web resources like MusicBrainz metadata, providing an enriched context for the audio content. For example, the files can be identified using the MusicDNS web service (<http://musicbrainz.org/doc/MusicDNS>) and thereby linked with all associated metadata available in the MusicBrainz database. The research workflow can be optimised through a SAWA facility that automatically stores feature extraction results so that these can be reused without having to perform the computation again.

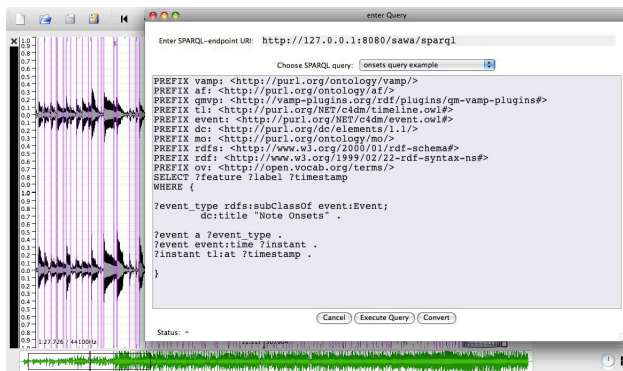


Fig. 3. SPARQL client in Sonic Visualiser

## 4 Conclusions

Our present work provides an overview of how the new Audio Feature Ontology in combination with related linked data vocabularies and open source research tools can be used to form a modular research framework for music information retrieval. The user is introduced to a simple Web-based interface as well as an open source desktop application, both of which enable extraction of musically meaningful information - such as pitch, loudness, timbre or note onsets - from audio files. The audio files can be linked up with online music metadata providing enriched content for music research workflows.

**Acknowledgements.** The development of the components presented in this demo was supported by the Engineering and Physical Sciences Research Council (EPSRC) through the OMRAS2 project EP/E017614/1, and by the JISC (<http://jisc.ac.uk>) funded Shared Open Vocabulary for Audio Research and Retrieval (SOVARR) project (<http://sovarr.c4dm.eecs.qmul.ac.uk/>).

## References

1. Fazekas, G., Raimond, Y., Jakobson, K., Sandler, M.: An overview of semantic web activities in the OMRAS2 project. *Journal of New Music Research (JNMR)* 39(4) (2010)
2. Raimond, Y., Abdallah, S.A., Sandler, M.B., Giasson, F.: The Music Ontology. In: *Proc. 8th International Conference on Music Information Retrieval*, Vienna, Austria, September 23-27 (2007)
3. Cannam, C., Landone, C., Sandler, M.B.: Sonic visualiser: an open source application for viewing, analysing, and annotating music audio files. In: Bimbo, A.D., Chang, S.-F., Smeulders, A.W.M. (eds.) *ACM Multimedia*, pp. 1467–1468. ACM (2010)

4. Fazekas, G., Cannam, C., Sandler, M.B.: Reusable Metadata and Software Components for Automatic Audio Analysis. In: Proc. IEEE/ACM Joint Conference on Digital Libraries (JCDL 2009) Workshop on Integrating Digital Library Content with Computational Tools and Services, Austin, Texas, USA (2009)
5. Page, K.R., Fields, B., Nagel, B.J., O'Neill, G., Roure, D.D., Crawford, T.: Semantics for music researchers: How country is my country? In: Polleres, A., Chen, H. (eds.) ISWC Posters & Demos. CEUR Workshop Proceedings, vol. 658. CEUR-WS.org (2010)