Ontological Representation of Audio Features

Alo Allik(✉), György Fazekas, and Mark Sandler

Queen Mary University of London, London, UK
{a.allik,g.fazekas,mark.sandler}@qmul.ac.uk

Abstract. Feature extraction algorithms in Music Informatics aim at deriving statistical and semantic information directly from audio signals. These may be ranging from energies in several frequency bands to musical information such as key, chords or rhythm. There is an increasing diversity and complexity of features and algorithms in this domain and applications call for a common structured representation to facilitate interoperability, reproducibility and machine interpretability. We propose a solution relying on Semantic Web technologies that is designed to serve a dual purpose (1) to represent computational workflows of audio features and (2) to provide a common structure for feature data to enable the use of Open Linked Data principles and technologies in Music Informatics. The Audio Feature Ontology is based on the analysis of existing tools and music informatics literature, which was instrumental in guiding the ontology engineering process. The ontology provides a descriptive framework for expressing different conceptualisations of the audio feature extraction domain and enables designing linked data formats for representing feature data. In this paper, we discuss important modelling decisions and introduce a harmonised ontology library consisting of modular interlinked ontologies that describe the different entities and activities involved in music creation, production and publishing.

Keywords: Semantic audio analysis · Music Information Retrieval · Linked open data · Semantic Web technologies

1 Introduction

The availability of unprecedented amounts of music in digital formats is dramatically changing the way casual and professional users interact with large music collections on the Web. Using textual editorial metadata is no longer sufficient and reliable as the principal means of finding the desired content. Statistical and musical information extracted from digital audio is becoming an increasingly valuable ingredient in strategies for searching, discovering and browsing music in large collections. These strategies are a result of intensive research and development in the Music Information Retrieval (MIR) community with active participation stemming from both academic and commercial interests. Consequently, there is a growing diversity of audio feature extraction algorithms combined with a profusion of audio feature datasets available for research communities and commercial developers. However, it is not always clear what certain

© Springer International Publishing AG 2016
P. Groth et al. (Eds.): ISWC 2016, Part II, LNCS 9982, pp. 3–11, 2016.
DOI: 10.1007/978-3-319-46547-0_1

feature data represents or why two extraction algorithms, identified as identical by their developers, may produce strikingly dissimilar results when applied to the same audio signal. The situation is exacerbated by the lack of common terminology or structuring principles in existing data interchange formats that often have a narrow scope to satisfy tool or task specific requirements. There is a need for more meaningful representation of feature data that would facilitate linking or comparing features produced in different data sources, as well as for generalised descriptions of audio features that would allow easier identification and comparison of audio feature algorithms that produce the data.

We propose a modular approach using Semantic Web ontologies for the representation of audio features. The Audio Feature Ontology framework consists of two separate components, *(i)* a core ontology and *(ii)* a separately maintained extensible vocabulary. This is motivated by the need for mediation between several tool and task specific conceptualisations that exist in this diverse domain. The Audio Feature Vocabulary includes existing audio features and captures computational workflows, providing the terms for specific ontologies without attempting to organise the features hierarchically. The Audio Feature Ontology represents entities in the feature extraction process on different levels of abstraction, modelling the underlying activities involved in problem solving through phases of conceptualisation, modelling and implementation.

2 Background

The need for an ontological representation of audio features was already recognised during the development of the Music Ontology framework [7]. This framework consists of a harmonised library of modular music-related ontologies [1] including a feature ontology. The early version of this ontology was primarily designed to provide terms for the Vamp plugin system[1], an extensible collection of feature extraction algorithms that accept audio signals as input and produce structured feature data as output, including formats prescribed by the original ontology. The plugins are executed in host applications such as the command line Sonic Annotator[2] tool and Sonic Visualiser[3], a desktop application designed to provide visualisations of audio feature data. A number of MIR libraries also release their feature extractors as Vamp plugins. This system has so far been the only solution enabling a shared ontologically structured data representation of audio features. However, the initial ontology does not provide a comprehensive vocabulary of audio features or computational feature extraction workflows. It also lacks concepts to support development of more specific feature extraction ontologies, while structurally it conflates musicological and computational concepts in a way that makes it inflexible for certain modelling requirements [2].

Other existing feature extraction frameworks provide data exchange formats designed for particular workflows or specific tools, providing interoperability on

[1] http://www.vamp-plugins.org.
[2] http://www.vamp-plugins.org/sonic-annotator/.
[3] http://www.sonicvisualiser.org.

the syntactic level. However, there is no common structuring principle shared by these different tools and libraries. The motley of output formats is well demonstrated in the representations category of a recent evaluation of feature extraction toolboxes [6]. For example, the popular MATLAB MIR Toolbox[4] export function outputs delimited files as well as Weka Attribute-Relation File Format (ARFF), while Essentia[5] provides YAML and JSON and the YAAFE library outputs CSV and HDF5. The MPEG-7 standard, used as benchmarks for other extraction tools, provides an XML schema for a set of low-level descriptors. The most recent developments in audio feature data formats predominantly employ JavaScript Object Notation (JSON), which is rapidly becoming a ubiquitous data interchange mechanism in a wide range of systems regardless of domain. It is evident that the simplicity of JSON combined with its structuring capabilities make it an attractive option, particularly compared to preceding alternatives including YAML, XML, ARFF, the Sound Description Interchange Format (SDIF) and various delimited formats.

While existing RDF-based solutions face criticism by some domain experts [3], suggesting they are non-obvious, verbose or confusing, we believe this should be addressed in the ontology engineering process. The potential of interoperable representation of audio features on the semantic rather than the syntactic level and the ability to link features with other music related information provides a more sustainable platform for researchers and commercial developers alike. This is in stark contrast with solutions that do not support linking through unique identification of entities existing at different conceptual levels, and do not publish their schema using standardised languages that allow formalising relations, not only concept hierarchies, in this complex domain.

3 Core Ontology Model

In order to address the issues of domain structuring and data representation, we propose a modular framework for the Audio Feature Ontology, separating abstract ontological concepts from more specific vocabulary terminology. The framework also provides for describing extraction workflows and increases flexibility for modelling task and tool specific ontologies. The core structure of the framework separates the underlying classes that represent abstract concepts in the domain from specific named entities. This results in the two main components of the framework defined as:

- Audio Feature Ontology (AFO): https://w3id.org/afo/onto/1.1#
- Audio Feature Vocabulary (AFV): https://w3id.org/afo/vocab/1.1#

The ontology component is structured to reflect different conceptual levels of abstraction of audio features. These layers represent the design process from *(i)* conceptualisation of a feature, through *(ii)* modelling an algorithmic workflow, to

[4] http://bit.ly/1rCwJOt.
[5] http://essentia.upf.edu.

(iii) implementation and *(iv)* instantiation in a specific computational context. For example, the abstract concept of Chromagram[6] is separate from its algorithmic model, which involves a sequence of computational operations like cutting an audio signal into frames, calculating the Discrete Fourier Transform for each frame, etc. (see Sect. 4 for a more detailed example). The computational workflow can be implemented in different ways and in various programming languages as components of feature extraction libraries. The implementation layer enables distinguishing a Chromagram written as a Vamp plugin from a Chromagram extractor in the MIR Toolbox. The most concrete layer represents the feature extraction instance in a specific execution context, for example, to reflect the differences of operating systems or hardware on which the extraction occurred. Our proposed layered model is shown in Fig. 1.

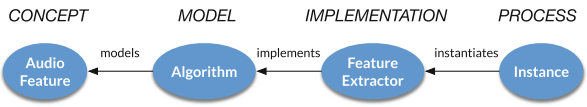

Fig. 1. The Audio Feature Ontology core model with four levels of abstraction

The core model of the ontology retains original attributes to distinguish audio features by temporal characteristics and data density. It relies on the Event[7] and Timeline[8] ontologies to provide primary structuring concepts for feature data representation. Temporal characteristics classify feature data either into instantaneous points in time - e.g. event onsets or tonal change moments - or events with known time duration. Data density attributes allow describing how a feature relates to the extent of an audio file: whether it is scattered and occurs irregularly over the course of the audio signal (for example, segmentation or onset features), or the feature is calculated at regular intervals and fixed duration (e.g. signal-like features with regular sampling rate). Figure 2 illustrates how audio features are linked with terms in the Music Ontology and thereby other music-related metadata on the Web. Specific named audio feature entities, such as **afv:Onset**, **afv:Key**, and **afv:MFCC** are subclasses of **afo:AudioFeature**, which, in turn, is a subclass of **event:Event** from the Event Ontology. This way the feature data can be directly linked to time points on the audio signal timeline using the **event:time** property. Listing 1.1 shows a Turtle/RDF example of such linking.

Since there are many different ways to structure audio features depending on a specific task or theoretically motivated organising principle, a common representation would have to account for multiple conceptualisations of the domain and facilitate diverging representations of common features. For example, Mel

[6] A feature representing energies of harmonically related frequencies calculated in discrete steps over time. Different musical temperaments yield different chromagrams.
[7] http://motools.sourceforge.net/event/.
[8] http://motools.sourceforge.net/timeline/.

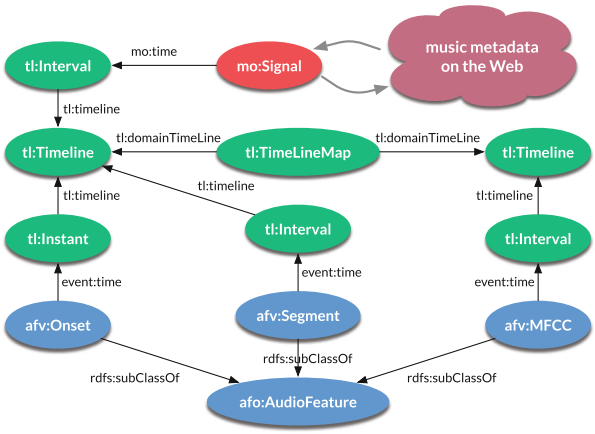

Fig. 2. Framework model showing how feature data representation is linked with music metadata resources on the Web using temporal entities defined in the Timeline ontology

Frequency Cepstral Coefficients (MFCC), that measure rates of energy change in different frequency bands and are widely calculated in many tools and workflows, can be categorised as a "timbral" feature in the psychoacoustic or musicological sense (as in MIR Toolbox for instance), while from the computational point of view, MFCCs could be labelled as a "cepstral" (e.g. in [5]) or "spectral" representation (as in the Essentia library). Collated audio features gathered from relevant literature and extraction software are defined as subclasses in the AFV. Another role of the vocabulary is to define computational extraction workflow descriptions, so that features can be more easily identified and compared by their respective computational signatures. This is discussed in the following section in more detail.

4 Algorithmic Workflow Representation

AFV defines terms that may be subsumed in specific ontologies and implements the *model* layer of the ontology framework. It is a clean version of the catalogue which lists the features without their properties. Many duplications of terms are consolidated. This enables the definition of tool and task specific feature implementations and leaves any categorisation or taxonomic organisation to be specified in the implementation layer.

The vocabulary also specifies computational workflow models for some of the features which lower-level ontologies can be link to. The computational workflow models are based on feature signatures as described in [5]. The signatures represent mathematical operations employed in the feature extraction process with each operation assigned a lexical symbol. It offers a compact description of each feature and facilitates the comparison of features by their computation

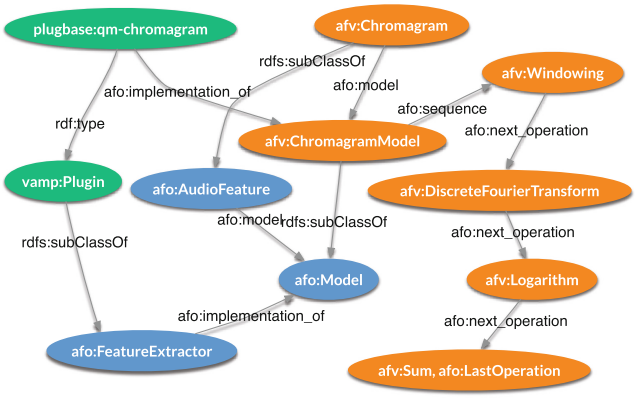

Fig. 3. Computational workflow of the Chromagram feature model linked to the extractor algorithm implemented in a Vamp plugin

workflows. The ontological representation of signatures involves defining a set of OWL classes that describe the representation and sequential nature of the calculations. The operations are implemented as sub-classes of three general classes: transformations, filters and aggregations. For each abstract feature, we define a model property. The OWL range of the model property is a ComputationalModel class in the Audio Feature Ontology namespace. The operation sequence can be defined through this object's operation sequence property. For example, the signature of the Chromagram feature is defined in [5] as "f F l Σ", which designates a sequence of (1) windowing (f), (2) Discrete Fourier Transform (F), (3) logarithm (l) and (4) sum (Σ). Figure 3 shows the resulting graph of the workflow.

5 Audio Content Description

Besides representing the computational steps involved in the extraction process, the framework supports identifying an extracted audio feature by linking it to a corresponding term in the Audio Feature Vocabulary, describing the temporal structure and density of the output data, associating feature data as intervals or instants on the audio signal timeline and associating the output data with feature extraction tools used in the extraction process. It also provides terms to represent inputs and parameters to the feature extraction functions to provide support for development of ontologies specific to a software library.

AFO can facilitate the development of other data formats beside RDF/Turtle that are aligned with linked data principles, such as JSON-LD [4]. JSON-LD is a linked data extension to the standard JSON format that provides an entity-centric representation of RDF/OWL semantics and a means to define a linked data context with URI connections to external ontologies and resources. It has

the potential to simplify feature representations while maintaining ontological structuring of the data.

Content-based analyses are becoming crucial in recommendation systems to tackle problems of rarely accessed content for which listening data supporting collaborative filtering is unavailable. These archives are important part of the Web and should be better represented and made accessible on the Semantic Web. The ontology is also a candidate to provide linked data representation for AcousticBrainz[9], which currently includes content-based metadata for over 2 million audio tracks. Adaptation in this context will facilitate significant deployment of musical metadata as linked data, where the feature identification and provenance data describing algorithms, computational tools and services are crucial for interoperability and wider utilisation of such data. The ontology has also been used in large-scale feature extraction projects such as Digital Music Lab[10] and Computational Analysis of the Live Music Archive[11]. The ontology can be deployed to describe large content-based music archives in libraries, music labels and open archives such as the Internet Archive Live Music Archive.

```
:track_1 a mo:Track ;
    dc:title"Afterlife";
    foaf:maker [ a mo:MusicArtist; foaf:name"Desimal"] ;
    mo:available_as <file:///snd/Afterlife.mp3> .

<file:///snd/Afterlife.mp3> a mo:AudioFile ;
    mo:encodes :signal_f6261475.

:signal_f6261475 a mo:Signal ;
    mo:time [
        a tl:Interval ;
        tl:onTimeLine :timeline_aec1cb82
    ] .

:timeline_aec1cb82 a tl:Timeline .

:event_1 a afv:Onset ;
    event:time [
        a tl:Instant ;
        tl:onTimeLine :timeline_aec1cb82 ;
        tl:at"PT1.98S"^^xsd:duration ;
    ] ;

:feature_1 a afv:MFCC ;
    mo:time [
        a tl:Interval ;
        tl:onTimeLine :timeline_aec1cb82 ;
    ] ;
    afo:value ( -26.9344 0.188319 0.106938 ..) .
```

Listing 1.1. An abbreviated example of linking onsets and MFCC features using AFV to the Music Ontology

Beyond representing audio feature data in research workflows, there are many other practical applications for the ontology framework. One of the test cases is providing data services for an adaptive music player that uses audio features to enrich user experience and enables novel ways to search or browse large music collections. The data is used by Semantic Web entities called Dynamic Music

[9] http://acousticbrainz.org.
[10] http://dml.city.ac.uk.
[11] http://etree.linkedmusic.org/about/calma.html.

Objects (dymos) [8] that control the audio mixing functionality of the player. Dymos make song selections and determine tempo alignment for cross-fading based on features.

6 Conclusions

The Audio Feature Ontology and Vocabulary provide a framework for representing the semantics of audio features providing interoperability on the conceptual rather than the syntactic level. It provides terminology to facilitate task and tool specific ontology development and serves as a descriptive framework for audio feature extraction. The proposed framework is a significant update to the existing ontology that addresses shortcomings of the original model, which have been identified as barriers to wider adoption in the community. The updates to the original ontology for audio features strive to simplify feature representations and make them more flexible while maintaining ontological structuring and linking capabilities. We produced example ontologies for existing tools including MIR Toolbox, Essentia, and Marsyas. Existing feature extraction tools, including the Sonic Visualiser and Sonic Annotator have been updated to produce RDF/Turtle as well as JSON-LD output. More examples of feature data representation, case studies of use of the ontology framework in emerging applications, and suggestions for best practices are available online: https://w3id.org/afo/onto/1.1#.

Acknowledgments. This work was supported by EPSRC Grant EP/ L019981/1, "Fusing Audio and Semantic Technologies for Intelligent Music Production and Consumption" and the European Commission H2020 research and innovation grant AudioCommons (688382). Sandler acknowledges the support of the Royal Society as a recipient of a Wolfson Research Merit Award.

References

1. Fazekas, G., Raimond, Y., Jakobson, K., Sandler, M.: An overview of semantic web activities in the OMRAS2 project. J. New Music Res. (JNMR) **39**(4), 295–311 (2010)
2. Fields, B., Page, K., De Roure, D., Crawford, T.: The segment ontology: bridging music-generic and domain-specific. In: Proceedings of the IEEE International Conference on Multimedia and Expo, 11–15 July 2011, Barcelona, Spain (2011)
3. Humphrey, E.J., Salamon, J., Nieto, O., Forsyth, J., Bittner, R., Bello, J.P.: JAMS: a JSON annotated music specification for reproducible MIR research. In: Proceedings of the 15th International Society for Music Information Retrieval Conference, Taipei, Taiwan (2014)
4. Lanthaler, M., Gütl, C.: On using JSON-LD to create evolvable RESTful services. In: Proceedings of the 3rd International Workshop on RESTful Design at WWW 2012 (2012)
5. Mitrovic, D., Zeppelzauer, M., Breiteneder, C.: Features for content-based audio retrieval. In: Advances in Computers, vol. 78, pp. 71–150 (2010)

6. Moffat, D., Ronan, D., Reiss, J.D.: An evaluation of audio feature extraction toolboxes. In: Proceedings of the 18th International Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway (2015)
7. Raimond, Y., Abdallah, S., Sandler, M., Giasson, F.: The music ontology. In: Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007), 23–27 September, Vienna, Austria (2007)
8. Thalmann, F., Carillo, A.P., Fazekas, G., Wiggins, G.A., Sandler, M.: The mobile audio ontology, experiencing dynamic music objects on mobile devices. In: Proceedings of the 10th IEEE International Conference on Semantic Computing, Laguna Hills, CA, USA (2016)