TOWARDS ONTOLOGICAL REPRESENTATIONS OF DIGITAL AUDIO EFFECTS

Thomas Wilmering, György Fazekas and Mark B. Sandler

Centre for Digital Music
Department of Electronic Engineering
Queen Mary, University of London
Mile End Road, London E1 4NS, UK
{thomas.wilmering, gyorgy.fazekas, mark.sandler}@eecs.qmul.ac.uk

ABSTRACT

In this paper we discuss the development of ontological representations of digital audio effects and provide a framework for the description of digital audio effects and audio effect transformations. After a brief account on our current research in the field of high-level semantics for music production using Semantic Web technologies, we detail how an Audio Effects Ontology can be used within the context of intelligent music production tools, as well as for musicological purposes. Furthermore, we discuss problems in the design of such an ontology arising from discipline-specific classifications, such as the need for encoding different taxonomical systems based on, for instance, implementation techniques or perceptual attributes of audio effects. Finally, we show how information about audio effect transformations is represented using Semantic Web technologies, the Resource Description framework (RDF) and retrieved using the SPARQL query language.

1. INTRODUCTION

Research in musical applications of metadata have produced a rich literature in recent years. This includes use cases for creative music production as well as information retrieval. Brazil [1] for example exploits *cue points* or markers, which can be found in modern audio file formats, for browsing music collections. Gomez [2] introduces the use of metadata for content-based audio processing, while Pampalk [3] uses metadata for organising sample libraries. The use of content-derived information for creating adaptive audio effects was described by Verfaille et al. [4]. Previously, the authors also exploited metadata in creative applications including navigation of recording projects using segmentation [5], and more recently we introduced a new class of audio effects where the use of standardised metadata is deeply embedded into the process of applying audio effects [6].

This system enables the prediction of changes in metadata when simple effects are applied to an audio signal, and also provides means for tracking the application of audio effects in the music production workflow. During the development of this system we identified the need for closely linked information describing data flow in different system components, and the need for a common way of representing information about audio features, as well as the characteristics and parameters of audio effects. These requirements point to the need for using a common knowledge representation framework for inter-disciplinary classification of audio effects. While such a classification has been proposed previously [7], standardised schema were not employed to represent this knowledge.

We opt for adopting Semantic Web [8] technologies for our purposes, in recognition that they provide a uniform way of encoding and linking information, governed by shared ontology schema, as well as support high-level logical reasoning based on Description Logics [9]. In particular, we use Semantic Web ontologies which provide for an *explicit specification of a conceptualisation* [10]. Data expressed using our ontologies support a wide range of use cases in creative music production, as well as exchanging accurate production data between tools, and sharing data for example with an artist community on the Semantic Web. In our research we exploit previous work on developing such ontologies and applications (see [11] for details) and develop an ontology based representation of audio effects within a common ontological framework for representing music related information and in particular studio production. This distinguishes our work from previous research where the use of metadata was only considered in isolation.

In the rest of this paper, after a brief review of Semantic Web technologies, we give an overview of the Music [12] and Studio Ontologies[1]. We discuss different approaches of developing an audio effects ontology, and demonstrate a use case of retrieving detailed information using metadata describing a musical mixture.

2. SEMANTIC WEB TECHNOLOGIES

Semantic Web Technologies refer to a set of web standards for creating a "Web of Data". The purpose of the Semantic Web, as an extension to the World Wide Web, is to allow for the development of applications that are capable of exploiting the meaning of knowledge represented in Web pages. At its core, Uniform Resource Identifiers (URI) are assigned to each resource including ontological concepts and relationships, while the Resource Description Framework (RDF) defines the standard for the formal description of these resources. The RDF data model expresses statements about resources as sets of *triples* in the form of *subject, predicate, object*. The model can be seen as directed graphs, where nodes represent the subjects and objects of statements while, arcs correspond to predicates[2]. Graph nodes may either be named by URIs, literals (e.g. strings or numbers), or *blank nodes*[3]. The example in listing 1 shows how an audio effect implementation is described in RDF. The triple `:rdfx_delay` `fx:implementation_of` `fx:Echo` identifies the subject as an implementation of an echo effect.[4] The following lines describe further attributes of the im-

[1] Available from: http://motools.sourceforge.net/
[2] Predicate describes relationships between subjects and objects
[3] Nodes may remain unlabelled for brevity.
[4] Namespace prefixes such as *fx* correspond to ontologies.

plementation, such as plugin type, name, and rights-related data using DCMI Metadata Terms[5]. This representation makes use of the RDF Schema Language (RDFS) for describing properties and classes of RDF resources, and our ontology developed in the Web Ontology Language (OWL)[6] for further refinements providing unambiguous representation of data and data relationships.

```
:rdfx_delay fx:implementation_of fx:Echo;
     fx:plugin_type fx:Rdfx;
     dc:description "Feedback Delay";
     dc:rights "Copyright (c) 2010-2011 QMUL";
     dc:title "RDFx_Delay_1";
     foaf:maker [ a foaf:Agent ;
                  foaf:name "Thomas Wilmering" ] .
```

Listing 1: RDF data describing an audio effect implementation.

3. THE MUSIC AND STUDIO ONTOLOGY FRAMEWORKS

Semantic Web technologies play an increasingly large role in information management for multimedia applications. Detailed ontologies dealing with various aspects of music related information have already been published for music information retrieval (MIR) use cases. The Music Ontology in particular defines concepts and relationships for this domain, on top of two fundamental ontologies: Event[7] and Timeline[8] for expressing time-based events. This can be used to represent concepts such as a *recording session*, but also, with further ontological support, concepts such as note onsets, or the extent of a key segment in an audio file [11]. Our research extends this modular framework with ontologies for metadata-generation and usage in the music production environment.

The Studio Ontology (STUDIO) describes recording studio concepts and provides a framework for collecting metadata in audio production. It provides extensions containing specialised terms, including an ontology of multitrack recording[9], which enables linking elements of multitrack production tools, for example audio clips and tracks, to more general Music Ontology data [13]. The Audio Effects Ontology (FXO) is developed within this framework. The application of audio effects is an integral part of contemporary music production, therefore it was included in the core Studio Ontology. However, the need for accommodating different view points in audio effect classification gave rise to its modularisation. For example, classification based on implementation techniques or perceptual attributes require different ontology modules. The problems arising when attempting to unify these different approaches have been discussed in [7], and an interdisciplinary classification system was proposed. We show that Semantic Web ontologies, as opposed to classic taxonomies, provide a way to describe audio effects not only for classification purposes, but also for the creation and retrieval of detailed metadata about a music production.

4. THE AUDIO EFFECTS ONTOLOGY

This section deals with metadata requirements for audio effects and the development of the Audio Effects Ontology consisting of three parts, one describing effect transformations, another providing a DAFX taxonomy and a third defining provenance terms. The motivation behind the development of the Audio Effects Ontology is to describe the domain of audio effects taking into account different view points. A first step towards this is the development of several ontologies each covering the perspective of a particular discipline, such as composition, post-production/audio engineering or effects development. In this chapter we focus on classification strategies for audio effects in ontological representations.

4.1. Effect Classification

As mentioned in §3 there are different schemata by which audio effects can be classified. For instance, we can group audio effects by the perceptual attributes that are mainly modified by their application. This classification system may be the most natural for a composer, who is primarily interested in the aesthetics of a particular sound transformation. Table 1 shows a selection of effects and the modified perceptual attributes [14][4][7]. For each effect main attributes are identified and additionally one or more other attributes that are modified by to a lesser extent. The perceptual attributes comprise of *loudness*, *duration and rhythm*, *pitch and harmony*, *timbre and quality*, and *space*. Listing 2 shows the echo effect class definition in the Audio Effects Ontology. The ontology contains a class `fx:Fx` representing the superclass for effect classification. Subclasses, such as `fx:SpatialFx` and `fx:LoudnessFx` , are linked to the main perceptual attributes with a restriction on the predicate `fx:main_attribute`. Subclasses of these classes in turn inherit these attributes, hence we only directly link the effect to secondary perceptual attributes using restrictions on the predicate `fx:other_attribute`. The perceptual attributes are defined as individuals of the class `fx:PerceptualAttribute`.

DAFx Name	Perceptual Attribute Main	Perceptual Attribute Other
Distance Change	S	L,T
Directivity	S	P,T
Echo	S	L
Granular Delay	S	L,D,P,T
Panning	S	
Reverberation	S	L,D,T
Rotary Speaker	S	P,T
Filter	T	L
Comb Filter	T	L,P
Equaliser	T	L
Ring Modulation	P,T	
Robotisation	P,T	L
Spectral Tremolo	L,T	D
Spectral Warping	T,P	L
Time Shuffling	L,D,P,T	
Vibrato	L,P	T,D

Table 1: *Selection of digital audio effects and affected perceptual attributes (L: loudness, D: duration and rhythm, P: pitch and harmony, T: timbre and quality, S: space)[7][14].*

A technical classification based on implementation techniques on the other hand is not as straight-forward. However, such a taxonomy would be helpful for the developer interested in the relationships of effects based on underlying digital signal processing (DSP) algorithms. Verfaille et al. [7] proposed a technical classification based on [15]:

[5]http://dublincore.org/documents/dcmi-terms/
[6]OWL Referenece: http://www.w3.org/TR/owl-ref/
[7]http://purl.org/NET/c4dm/event.owl/
[8]http://purl.org/NET/c4dm/timeline.owl/
[9]The Multitrack Ontology http://purl.org/ontology/studio/multitrack

- filters
- delays (resampling)
- modulators and demodulators
- nonlinear processing
- spatial effects
- time-segment processing
- time-frequency processing
- source-filter processing
- spectral processing
- time and frequency warping

```
fx:SpatialFx a owl:Class ;
        rdfs:subClassOf fx:Fx ,
             [ rdf:type owl:Restriction ;
               owl:onProperty fx:main_attribute ;
               owl:hasValue fx:Spatial
             ] .

fx:Echo a owl:Class ;
        rdfs:subClassOf fx:SpatialFx ,
             [ rdf:type owl:Restriction ;
               owl:onProperty fx:other_attribute ;
               owl:hasValue fx:Loudness
             ] .
```

Listing 2: Description of the echo effect in the Audio Effects Ontology (perceptual).

Naturally, this type of classification has limits; some audio effects, e.g. pitch shifting, can be implemented by different techniques, thus some ambiguity is inherent in this system. Furthermore, one can argue that *spatial effect* is not an implementation technique as such, as it relates primarily to a modified perceptual attribute. In order to provide a detailed ontology for the technical description of audio effects a DSP ontology is desirable, and constitutes future work in this field of research.

In addition to the classification systems described above, we propose the development of a taxonomy from an audio engineering point of view. Here, it is important to clearly define the meaning of the term "audio effect". While in our research we mostly use the term in its general sense, equating the terms *audio effect* and *sound transformation*, from an audio engineer's point of view often a distinction is made between *effects* and *processors*. Such a classification system could be seen as lying in between a perceptual and technical system, based on the audio effects' roles in the production workflow. In audio engineering *effects* are more artistic in nature, altering the sound in a dramatic way, for instance a tapped delay or a flanger. Audio *processors* on the other hand are transformations aimed at the enhancement of sound mostly in post-production and mastering. Equalisers and certain compressors fall into this category. The online database for DAFx plugins by *KVR Audio*[10] lists *Mastering* as a separate category, containing mostly non-linear effects, such as enhancers and compressors designed to be applied on a musical mixture. A future direction of this research is a system unifying the audio effect ontologies of different disciplines, thus enabling and improving communication between developers, composers and audio engineers, providing a basis for the development of software agents processing audio effects related information to assist interdisciplinary work. Although we want to describe digital audio effects, we consider an effect as such as an acoustical phenomenon (e.g. an *Echo* is a series of reflections of a sound) which can then be linked to an implementation with its respective algorithm as its physical manifestation in order to describe audio effects software or transformations. This approach also distinguishes the ontology from the LV2 (Linux Audio Developers Simple Plugin API version 2) specification[11], which, although also written in RDF and containing a classification scheme, is limited to the description of effects implemented in LV2, without discerning implementations and audio effects as physical phenomena.

4.2. Effect Transformations

The Audio Effects Ontology also defines concepts for describing the application of effects to a signal. The class `fx:Transform` is used here, comparable to the way the Vamp Ontology[12] defines concepts for the application of feature extraction plugins. An `fx:Transform` may be linked to an effect implementation using `fx:Implementation`. This implementation class may also act as the connection to the taxonomy by linking it to `fx:Fx` with the property `fx:implementation_of` (an inverse property linking an effect type to an implementation is also given by `fx:implementation`). Using this system we are able to associate events on the audio signal timeline as defined by the Music Ontology to a particular transform, which in turn is associated to information about the implementation, the effect type, and the modified perceptual attributes. Moreover, we may also describe an adaptive effect implementation that processes metadata in the form of RDF data, e.g. note onsets described with concepts from the Audio Features Ontology[13][7][6]. We may describe the parameters of a sound effect implementation, both, to describe the settings at a particular transform, and to describe the available parameters for a given effect implementation. A set of standard parameters, such as the *dry/wet mix*, *feedback amount* and *gain* are defined in the ontology as well, which may be linked from an implementation parameter. In summary, the FXO is capable of describing audio effects in taxonomical systems adapted to different disciplines, effect implementations (e.g. plugins and hardware devices), and the application of audio transformations to audio signals.

5. QUERYING METADATA

Semantic metadata in RDF makes it possible to perform complex queries over the data using a query language such as SPARQL[14], assuming the metadata is accurately accumulated during the production process. Listing 3 shows RDF data describing an onset event on a signal timeline created by the application of an echo effect, in this case "RDFx_Delay_1". The description includes the parameter settings of the effect, as well as provenance data about the origin of the audio event in the multitrack project during production. Running the SPARQL query from listing 4 over the metadata associated with the resulting mixture returns all events created by audio effects modulating loudness as "other" perceptual attribute as shown in table 1. We may also query for additional information, such as plugin name or developer. The example shows that with this system we can retrieve information from a musical mixture otherwise lost during the mixing process, or only

[10] http://www.kvraudio.com/get.php

[11] http://lv2plug.in/ns/lv2core

[12] http://www.omras2.org/VampOntology/

[13] http://motools.sourceforge.net/doc/audio_features.html/

[14] http://www.w3.org/TR/rdf-sparql-query/

```
:transform_0 a fx:Transform;
fx:parameter_set [ fx:identifier "dryWetMix" ;
    fx:value "50"^^xsd:float ] ,
  [ fx:identifier "delayTime" ;
    fx:value "0.297"^^ xsd:float ] ,
  [ fx:identifier "feedback";
    fx:value "0"^^xsd:float ] ;
  fx:transform :rdfx_delay.

:event_0 a af:Onset ;
  event:time [ a tl:Instant ;
   tl:at "PT1.897007S"^^xsd:duration ;
   tl:onTimeLine :signal_timeline_0 ] ;
   fx:created_by_fx :transform_0 ;
   fx:track_origin :GuitarTrack .

:example_multitrack a mt:MultitrackProject ;
   mt:track :DrumTrack, :GuitarTrack .
```

Listing 3: RDF data describing an onset event on a signal timeline created by the application of the effect "RDFx_Delay_1".

retrievable by means of feature extraction from source-separated signals. However, tracking metadata throughout the production process makes source separation redundant in this context. Moreover, feature extraction from noisy audio material may pose problems concerning accuracy (this has been shown for the case of reverberation) [16].

```
SELECT ?time ?track WHERE {
       ?a event:time ?b ;
         fx:created_by_fx ?c ;
         fx:track_origin ?track.
       ?b tl:at ?time .
       ?c fx:transform ?d .
       ?d fx:implementation_of ?e .
       ?e fx:other_attribute fx:Loudness .
       }
```

Listing 4: SPARQL query retrieving all events created by a transformation modulating *loudness* as the secondary perceptual attribute. *Track* represents the original audio track in the multitrack production prior to the mixdown.

By integrating a query engine in an audio production system processing metadata according to the proposed ontology, the user is not only capable to select audio effects by semantic descriptors of different domains, but is also able to retrieve detailed metadata about workflows and techniques concerning audio productions.

6. CONCLUSIONS

In this paper we presented an ontology defining concepts for the application and classification of digital audio effects. We showed how this novel ontology fits in to existing Semantic Web ontologies, particularly the Music Ontology and the Studio Ontology extension. We showed that RDF metadata accumulated during the production process using the Studio Ontology Framework allows for the retrieval of detailed information about a music piece. The information may be used for musicological purposes revealing production workflows and the origin of individual audio events, or for further processing by passing it to content-based (adaptive) audio effects or feature extractors.

Our investigation of audio effects with respect to classification and ontology design showed that it is necessary to create multiple ontologies covering the different disciplines concerned (e.g. classifications based on implementation for developers or on perceptual attributes for composers). The presented work reveals the need for more specialised ontologies for our music information management framework, such as a dedicated signal processing ontology. Future work includes the development of such ontologies and further development and integration of the developed tools in music production applications.

7. REFERENCES

[1] E. Brazil, "Cue point processing: An introduction," *Proceedings of the COST G-6 Conference on Digital Audio Effects (DAFX-01), Limerick, Ireland, December 6-8*, 2001.

[2] E. Gomez, G. Peterschmitt, X. Amatriain, and P. Herrera, "Content-based melodic transformations of audio material for a music processing application," *in Proc. of the 6th Int. Conference on Digital Audio Effects (DAFx-03), London, UK, September 8-11*, 2003.

[3] E. Pampalk, P. Hlavac, and P. Herrera, "Hierarchical organization and visualization of drum sample libraries," *in Proc. of the 7th Int. Conference on Digital Audio Effects (DAFx-04), Naples, Italy, October 5-8*, 2004.

[4] V. Verfaille, U. Zölzer, and D. Arfib, "Adaptive digital audio effects (A-DAFx): A new class of sound transformations," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 5, 2006.

[5] G. Fazekas and M. Sandler, "Intelligent editing of studio recordings with the help of automatic music structure extraction," *presented at the AES 122nd Convention, Vienna, Austria*, 2007.

[6] T. Wilmering and M. Sandler, "RDFx: Audio effects utilising musical metadata," *presented at the 4th IEEE International Conference on Semantic Computing, Pittsburgh, PA, USA*, 2010.

[7] V. Verfaille, C. Guastavino, and C. Traube, "An interdisciplinary approach to audio effect classification," *Proceedings of the 9th International Conference on Digital Audio Effects (DAFx-06), Montreal, Canada*, September 2006.

[8] T. Berners-Lee, J. Hendler, and O. Lassila, "The semantic web," *Scientific American*, May 2001.

[9] I. Horrocks, "Ontologies and the Semantic Web," *Communications of the ACM*, vol. Vol. 51, no. 12, pp. pp. 58–67, 2008.

[10] T. R. Gruber, "A translation approach to portable ontology specification," *Knowledge Acquisition*, vol. 5, no. 2, 1993.

[11] G. Fazekas, Y. Raimond, K. Jakobson, and M. Sandler, "An overview of Semantic Web activities in the OMRAS2 Project," *Journal of New Music Research special issue on Music Informatics and the OMRAS2 Project*, vol. Vol. 39 issue 4, pp. pp. 295–311, 2010.

[12] Y. Raimond, S. A. Abdallah, M. Sandler, and F. Giasson, "The music ontology," *Proceedings of the International Conference on Music Information Retrieval*, 2007.

[13] G. Fazekas, Y. Raimond, and M. Sandler, "A framework for producing rich musical metadata in creative music production," *presented at the AES 125th Convention, San Francisco, CA, USA*, 2008.

[14] X. Amatriain, J. Bonada, À. Loscos, J. L. Arcos, and V. Verfaille, "Content-based transformations," *Journal of New Music Research*, vol. 32, no. 1, pp. 95–114, 2003.

[15] U. Zölzer, *DAFX - Digital Audio Effects*, J. Wiley & Sons, 2002.

[16] T. Wilmering, G. Fazekas, and M. Sandler, "The effects of reverberation on onset detection tasks," *presented at the AES 128th Convention, London, UK*, 2010.