Deliverable D4.6

Release of tool for the manual annotation of musical content

Grant agreement nr	688382
Project full title	Audio Commons: An Ecosystem for Creative Reuse of Audio Content
Project acronym	AudioCommons
Project duration	36 Months (February 2016 - January 2019)
Work package	WP4
Due date	29 September 2017 (M20)
Submission date	29 September 2017 (M20)
Report availability	Public (X), Confidential ( )
Deliverable type	Report ( ), Demonstrator ( ), Other (X)
Task leader	MTG-UPF
Authors	Xavier Favory, Eduardo Fonseca, Frederic Font
Document status	Draft (), Final (X)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Table of contents

Table of contents **2**

Executive Summary **3**

1 Annotation of musical content **4**

1.1 Current tools and interfaces for the annotation of musical content 4

1.1.1 Jamendo 4

1.1.2 Freesound 5

1.1.3 Freesound Datasets event-based annotation task 6

1.1.4. Discussion 7

1.2 Nomenclature/taxonomy - Important aspects to annotate 8

2 The Audio Commons Music Annotator **9**

2.1 Concept and main functionalities 9

2.2 Implementation 11

2.2.1 Content type and musical properties 11

2.2.2 Integration into web applications 12

3 Conclusion and Future Improvements **13**

4 References **14**

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Executive Summary

This deliverable presents the first prototype of a tool for the manual annotation of musical content present in the Audio Commons Ecosystem. This tool is a web-based interface that intelligently guides users on the annotation process of music sample and music pieces. The tool is released under an Apache 2.0 license and available at this public Github repository: https://github.com/AudioCommons/ac-annotator.

One of the challenges in making use of Creative Commons audio content comes from the fact that it is provided by various sources and authors with different backgrounds and levels of expertise. Therefore, the content is often unstructured and not properly annotated, which hinders its efficient retrieval. Moreover, there is a scarcity of tools and agreed methods to help users with the task of annotating musical content through established common procedures. Intelligently guiding users on the annotation process would allow a reliable and uniform description of the content which will therefore facilitate its sharing.

In this document, we first present the annotation tools used by some existing platforms of the Audio Commons Ecosystem, and then focus on a terminology of relevant musical concepts to be considered when manually annotating musical content. Then, we present the current prototype which is based on a two stage approach. First, the user identifies the type of musical content she wants to upload, and then selects musical attributes from a set of properties relevant for the selected content type. The outcome is a tool for the manual annotation of musical content that intelligently guides the user during the procedure using content-specific musical properties. We end this report with a summary of the work done and sketch the next steps to be carried out for the progressive improvement of the version presented here.

This deliverable will be complemented by Task 5.4, where an intelligent interface for manually annotating non-musical audio content will be developed. After integrating both tools, a combined evaluation will be performed leveraging real users.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

1 Annotation of musical content

One of the Audio Commons goals is to make audio content more accessible by providing an ecosystem where content can be easily shared and retrieved among a community of creative audio professionals. More specifically, we aim at facilitating the use of user-generated Creative Commons audio content such as sound effects, music recordings and music pieces.

One of the challenges in making use of this type of content comes from the fact that it is provided by various sources and authors with different backgrounds and levels of expertise. Therefore, the content is often unstructured and not properly annotated, which hinders its efficient retrieval. Moreover, there is a scarcity of tools and agreed methods to help users with the task of annotating musical content through established common procedures. Intelligently guiding users on the annotation process would allow a reliable and uniform description of the content which will therefore facilitate its sharing.

As part of the development of the Audio Commons Ontology, a sound schema is being developed which specifies a number of musical properties. Some of them have been included in the automatic annotation tool described in the deliverable D4.2 First prototype tool for the automatic semantic description of music samples. The schema proposes a differentiation between music samples and music pieces which tailor some descriptors to the particular context in which they are used. We define music pieces as audio recordings typically corresponding to complete songs, while music samples refers to individual or simpler music elements such as single notes, percussive hits, chords, melodies, chord progressions, rhythmic patterns or loops.

In this Section, we first analyse the annotation tools used by some platforms, and then focus on nomenclatures of relevant concepts that should be taken into account when manually annotating musical content.

1.1 Current tools and interfaces for the annotation of musical content

1.1.1 Jamendo

When users upload a music piece in Jamendo,[1] they are asked to fill different kinds of information:

- General information such as title of the piece, name of the album, etc…
- Lyrics
- Artwork
- Credits
- Tags and Metadata
- License

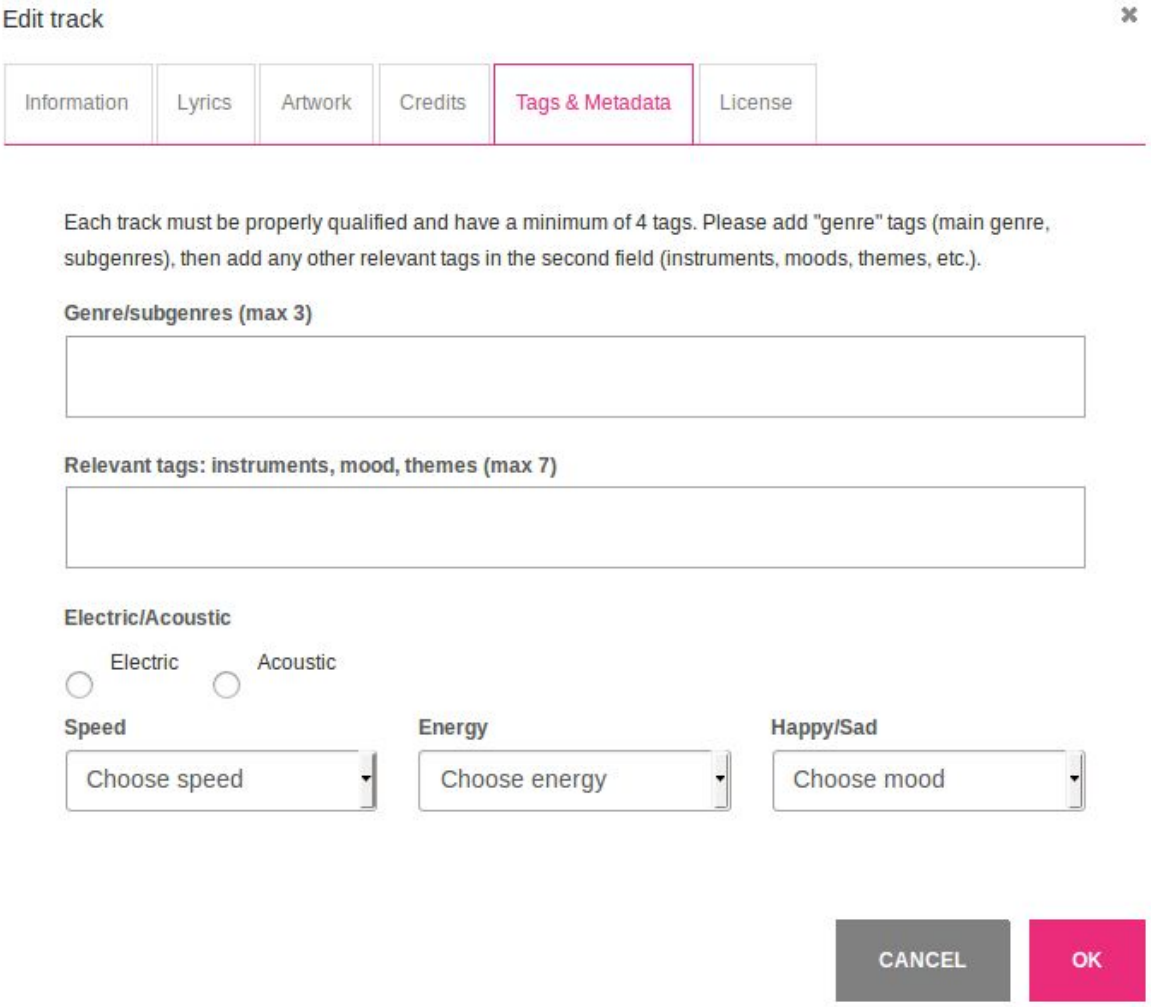
Among these, the category *Tags and metadata* is particularly relevant for facilitating the organization, access and retrieval of the content. It includes the following fields:

- Music genre from a set of predefined categories (maximum 3)
- Tags referring to instruments, theme, coloration, etc… (minimum 4 - maximum 7)
- Selection between Electric/Acoustic
- Speed (scale of 5 degrees)
- Energy (scale of 5 degrees)

[1] https://www.jamendo.com/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

- Happiness/Sadness (scale of 5 degrees)



Screenshot of the Jamendo annotation interface for Tags and Metadata

1.1.2 Freesound

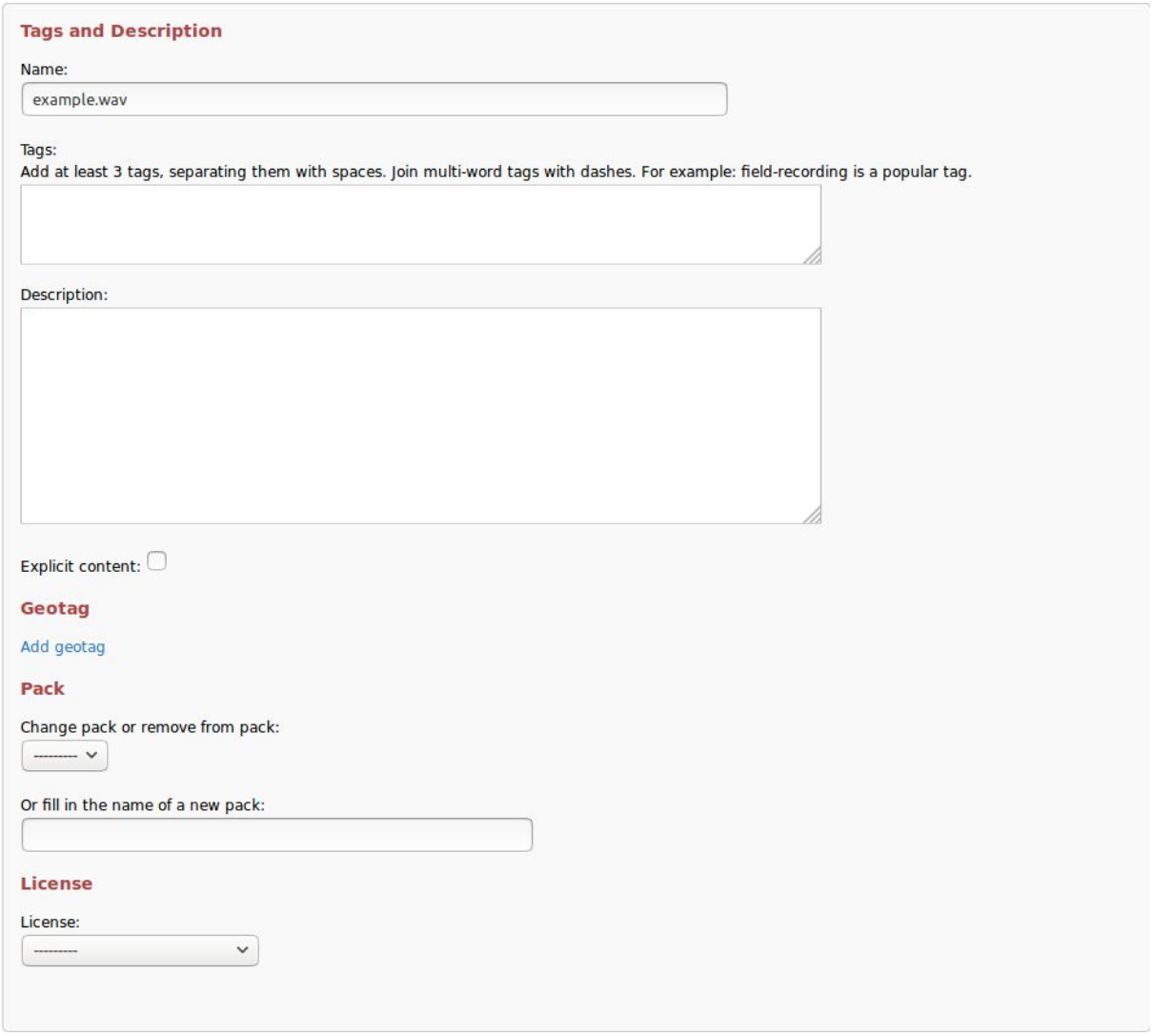
When sharing a sound in Freesound,[2] a user follows a description process where the following fields must be specified:

- Name of the sound
- Tags (minimum 3)
- Textual description
- Geotag (optional)
- Parental Advisory Explicit Content
- Assigned to a sound pack (optional)

[2] http://freesound.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

- License



Screenshot of the Freesound description interface

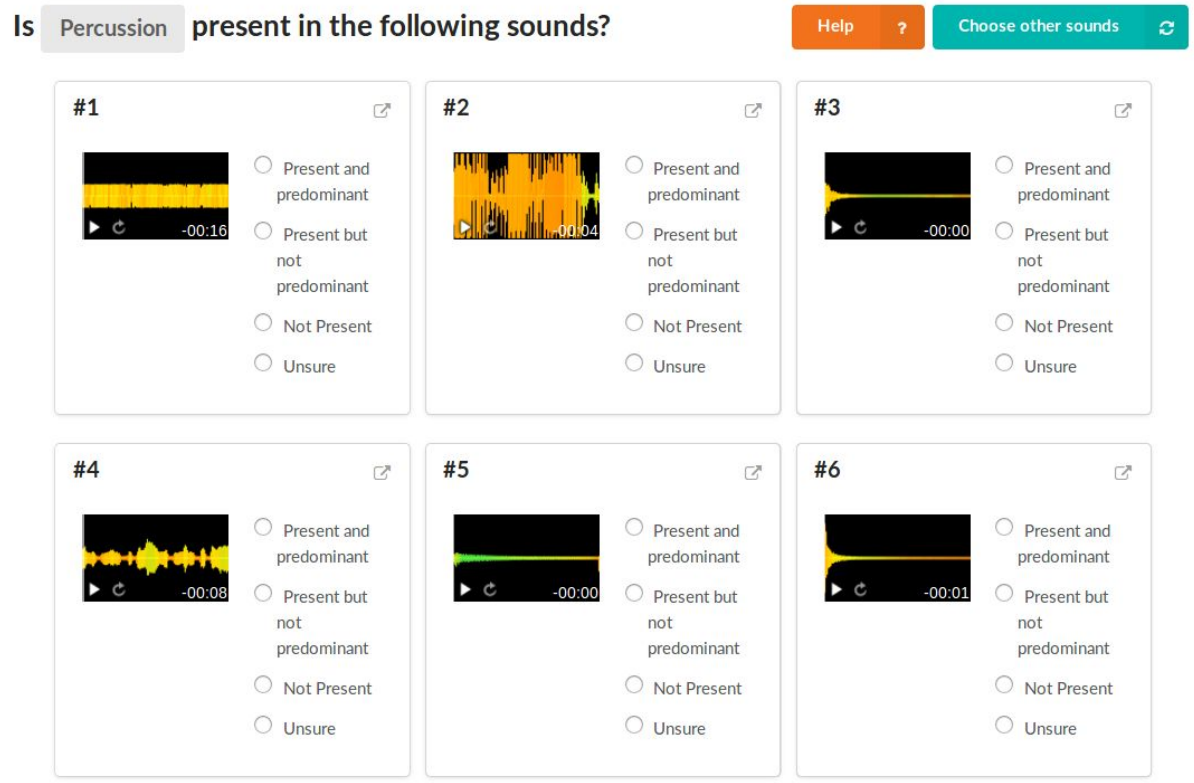
1.1.3 Freesound Datasets event-based annotation task

Freesound Datasets[3] is an online platform for the collaborative creation of open audio datasets [Fonseca17]. The platform is currently under development and at the moment only one annotation process is available. Unlike the aforementioned cases of Jamendo and Freesound where users provide information through a metadata *generation* task, in this case a *validation* task is carried out. It consists of the validation of previously generated annotations, where users are presented with a number of audio samples and, for each sample, they have to assess the presence of a particular sound category.

[3] https://datasets.freesound.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

This platform is being used for the creation of the FreesoundDataset (FSD), a general-purpose and large-scale audio dataset. It contains audio samples from Freesound labeled using the same hierarchical ontology of AudioSet, which includes 632 audio classes [Gemmeke17].



Screenshot of the Freesound Datasets Validate category annotations interface

1.1.4. Discussion

The validation task supported by the Freesound Datasets platform is considered more suitable as a post-annotation step, where users of a platform can collaboratively contribute, for instance, to ground truth generation. However, it is considered less appropriate for the process of publishing audio content in the Audio Commons Ecosystem, e.g., when content creators upload content to Freesound or Jamendo. Nevertheless, the Audio Commons Ecosystem benefits from Freesound Datasets as generated ground truth will be used to train models for annotation tools.

Leaving the validation task apart, the other reviewed interfaces for the manual annotation of musical content present the following aspects. They either focus on musical pieces (Jamendo) or on sound samples (Freesound), but they do not cover both types of content simultaneously. In the specific case of Freesound, all music samples are treated equally, that is, no distinction whatsoever is considered despite the heterogeneity inherent to this musical content. For instance, we argue that the aspects that happen to be relevant for musical loops may differ from those appropriate to describe a musical note or a sound texture. Should such a distinction be made, the annotation procedure could provide guidance to promote more uniform and consistent manual annotations, allowing ad hoc sets of attributes for determined musical content types.

To provide more complete descriptions, an approach in which users specify attributes of a given audio content is more appropriate. Therefore, we consider it more suitable for the process of

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

publishing audio content in the Audio Commons Ecosystem (e.g., when content creators upload content to Freesound or Jamendo). Furthermore, by taking inspiration from Ontology-based systems, i.e., taking advantage of the concepts' organization and their relations, it is possible to guide users in the annotation process. For instance, a set of relevant properties can be inferred from the specific musical content type. To give an example, a *musical note* should be attributed a note value whereas a more complex content like *musical loop* should be categorised with *genre* and *mood* concepts. By easing the task, we expect the annotation effort to be reduced. Also, by using content-specific relevant attributes, more uniform and consistent annotations will be gathered for improved organization and retrieval.

1.2 Nomenclature/taxonomy - Important aspects to annotate

Informed by the survey on user requirements[4] in D2.1 Requirements report and use cases, a first draft of the Audio Commons Ontology is currently under development. After reviewing the schema[5] proposed as the minimum requirement for a provider to support within the Audio Commons Ecosystem, and also taking into account other taxonomies such as that of Google's AudioSet Ontology,[6] we propose a list of content type categories for music samples and pieces to be considered in the manual annotation process. We also highlight the musical properties that are relevant to consider when annotating the aforementioned music resources.

Content Type	Relevant musical properties
Note	Instrument, note
Melody	Instrument, tonality, tempo, mood, genre
Chord	Instrument, chord
Chord progression	Instrument, tonality, tempo, mood, genre
Percussive hit	Percussion instrument
Rhythm pattern	Rhythm signature, tempo, mood, genre
Musical loop	Instruments, Tonality, tempo, signature, mood, genre
Texture/drone	Mood, genre
Music piece	Tonality, rhythm signature, tempo, mood, genre

List of the content type categories with their associated relevant musical properties. Rows 1-8 illustrate the content types defined for music samples. Music piece is considered a content type in itself.

[4]Audio Commons Initiative Survey on Creative Interaction with Audio Content: https://docs.google.com/forms/d/1c3iJhZcAPCqfmlUekj3aHAYPxeY6g-KW2bxZa8tUgNA/edit#responses
[5] https://docs.google.com/spreadsheets/d/19GlELgsx5AReb6d-8_6wfsVdSfUwCDggr8k2AXLVP48
[6] https://research.google.com/audioset/ontology/index.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

2 The Audio Commons Music Annotator

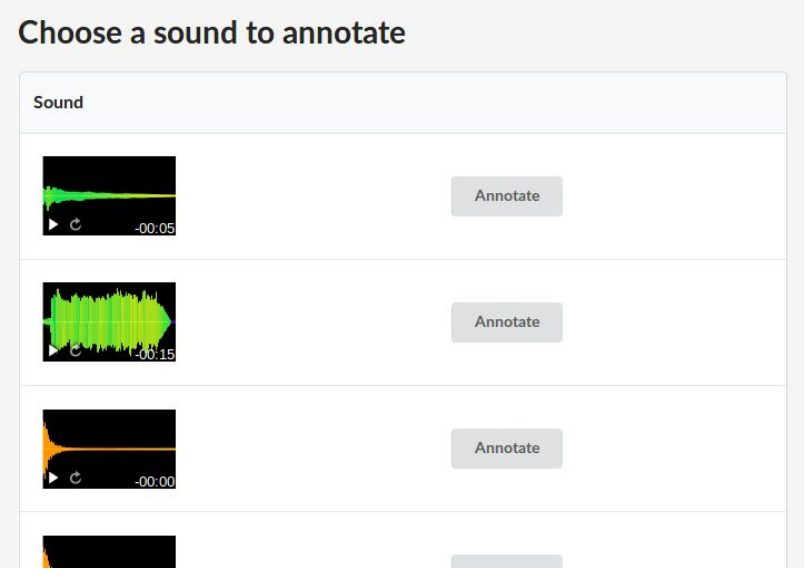
The Audio Commons Music Annotator is a web based tool for the manual annotation of musical content. It provides an interface which intelligently guides the user in the process of annotating music samples and music pieces. The interface developed is mainly thought for being used during the process of publishing audio content in the Audio Commons Ecosystem (e.g., when content creators upload content to Freesound or Jamendo). Nevertheless, it can also be used as a post-annotation step in which users of a platform can collaboratively contribute to the annotation of content and ground truth generation.

The prototype developed is not integrated in any content provider platform yet, but it is implemented in a way that it can be easily integrated in them. To show the functionalities and annotation flow of the prototype, we leverage a few music samples from Freesound.

2.1 Concept and main functionalities

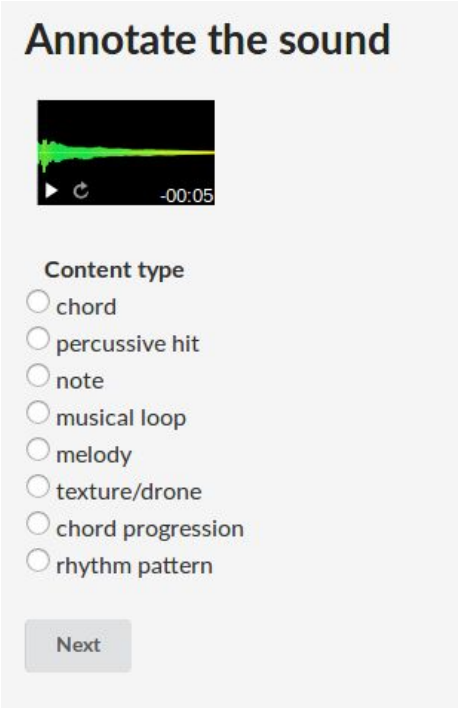
The main idea of the interface is to intelligently guide the user through the process of annotating audio content. For a given audio sample, following a two stage approach, the user first chooses a content type from a list of predefined categories. Then the user focuses on annotating musical properties associated to the content type previously selected.

1. In the developed prototype, the user first chooses a sound to annotate from a list of sounds:



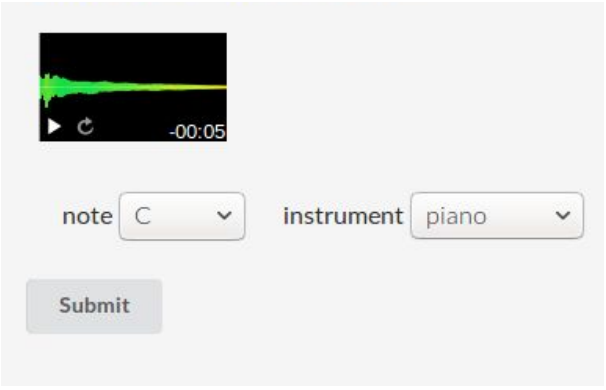
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

2. Then, the user chooses one of the proposed content types and moves forward to the next step:



3. According to the selected content type, different content-specific musical properties are presented:

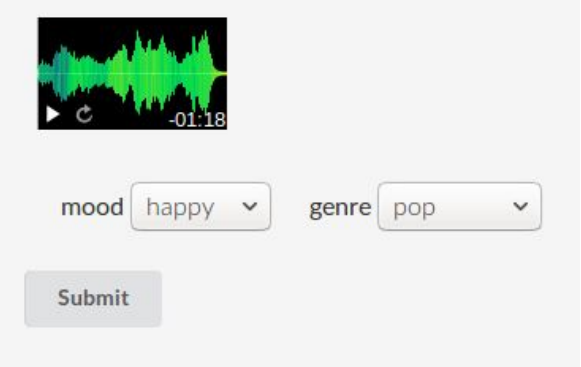
Note:



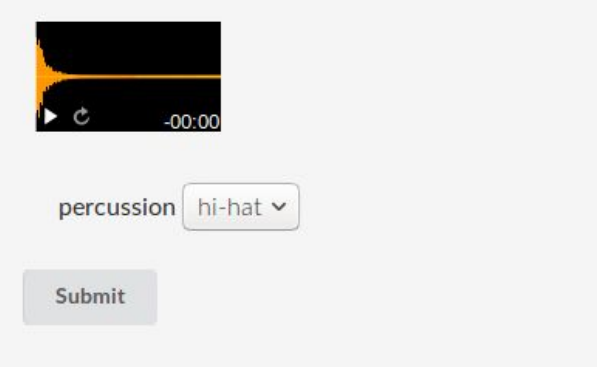
Melody:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

Texture/drone:



Percussive hit:



2.2 Implementation

The Audio Commons Annotator prototype demonstrator is implemented using the Python programming language and the popular Django[7] web framework. The source code is open source, released under Apache 2 license and available at https://github.com/AudioCommons/ac-annotator.

The Music Annotator is an application inside the ac-annotator Django project. The dynamic form presented is implemented in Javascript and takes as input a JSON file with the sound schema where the content type categories and the associated musical properties are specified. The sound schema is currently *hardcoded* in the JSON file, but in future iterations of the tool it will be loaded from the Audio Commons Ontology.

The easiest way to set up a local development version is to use Docker[8] container definitions provided in this repository. The instructions for setting up the development environment using Docker are the following:

- Before starting make sure you have Docker (with `docker-compose`) installed.
- Clone repository and cd into it:
  >>> git clone git@github.com:AudioCommons/ac-annotator.git
  >>> cd ac_annotator
- Build and run Docker containers for required services:
  >>> docker-compose up
- Now you should be able to access your server at `http://localhost:8000`

2.2.1 Content type and musical properties

As there are two steps in the annotation process, there are two main concepts that the dynamic form relies on: *content type* and *properties*.

As an example, here is a schema that the annotator can use as input:

[7] https://www.djangoproject.com/
[8] https://www.docker.com/products/overview

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

```
schema_dict = {
        "content_types": {
                "note":                 ["note", "instrument"],
                "rhythm pattern":       ["mood", "genre"],
                "chord progression":    ["mood", "genre"],
                "chord":                ["chord", "instrument"],
                "musical loop":         ["mood", "genre"],
                "percussive hit":       ["percussion"],
                "melody":               ["instrument", "mood", "genre"],
                "texture/drone":        ["mood", "genre"]
        },
        "properties": {
                "chord":        ["C", "C#/D ♭ ", "D", "D#/E ♭ ", "E", "F", "F#/G ♭ ", "G", "G#/A ♭ ", "A", "A#/B ♭ ", "B"],
                "mood":         ["happy", "funny", "sad", "tender", "exciting", "angry", "scary"],
                "percussion":   ["kick", "snare", "hi-hat", "tom", "crash", "ride"],
                "note":         ["C", "C#/D ♭ ", "D", "D#/E ♭ ", "E", "F", "F#/G ♭ ", "G", "G#/A ♭ ", "A", "A#/B ♭ ", "B"],
                "instrument":   ["piano", "guitar", "violin", "bass", "accordion", "saxophone", "trumpet"],
                "genre":        ["pop", "hip hop", "rock", "blues", "soul", "reggae", "country", "funk", "folk", "jazz",
                                "classical",  "electronic"]
        }
}
```

The *keys* in 'content_types' correspond to the different content types a user can specify (in this example, only types corresponding to musical samples are specified). The lists available for each of these content types include the musical properties that are relevant to be annotated for the specific content type. The schema showed here is only a temporary placeholder and presents just a few musical properties among all the important ones.

The keys in 'properties' refer to the different properties which are used for describing the different types of content. The possible values or attributes are listed for each property.

2.2.2 Integration into web applications

In this prototype, a Django template is used to create the HTML page that contains the main elements for allowing the Javascript to dynamically add the form.

The annotate function from `ac-annotator/music_annotator/views.py` sends the schema within the context of the template `ac-annotator/music_annotator/templates/annotate.html`. The HTML page rendered loads the Javascript script located in `ac-annotator/static/js/create_form.js` which uses the schema for creating the different forms.

The HTML template can be easily adapted to be used within another web development framework as it does not rely on Django related functionality, except for the Templates approach to get the schema from the server. The Javascript script can be used in any other environment, by taking care of defining the variable 'schema' (defined in the HTML template in this example).

When submitting the annotations, the results are formatted in a JSON fashion (e.g. { content_type: "chord progression", mood: "happy", genre: "pop" }). In this prototype, the results are logged on the client browser console. An AJAX POST request could be used to send the results as a JSON to the server.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

3 Conclusion and Future Improvements

In this deliverable, after a brief analysis of some interfaces for the manual annotation of musical content, we have described an initial version of a prototype tool for the manual annotation of musical samples and pieces. At the moment, the prototype consists of a web-based interface that intelligently guides users on the annotation process. It is based on a two stage approach and leverages content-specific musical properties.

The prototype developed here will continue evolving during the next phases of the project. Our plan is to progressively incorporate new features such as the inclusion of content-based recommendation of musical properties. More specifically, we intend to integrate the tool for automatic semantic description of music samples presented in Deliverable D4.2, with the aim of pre-filling some musical properties as a suggestion to the content creator. Moreover, we will consider allowing users to use their own words for specifying some of the musical properties, thereby allowing users to provide musical attributes beyond those available in the current closed-form. This would be useful for the cases where the musical content's particularities do not fit well within the attributes established in the current prototype version. Finally, for allowing full compatibility with the Audio Commons Ecosystem, the schema used for describing the properties that are relevant to be annotated for the specific content type, which serves here as an example, will be replaced and will take informations from the Audio Commons Ontology once fully accomplished.

The core of the current prototype will be reused for the development of Task 5.4 Research, development and evaluation of an interface for manually annotating non-musical content. As part of this task, an evaluation of the interface will be carried out in terms of its usability and its expressive power for annotating music samples and music pieces. This evaluation will be carried out with real users and in combination with the evaluation of the tool developed in Task 5.4.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N° 688382

4 References

[Fonseca17] Fonseca, Eduardo et al. (2017). “Freesound Datasets: A platform for the creation of open audio datasets”. In: Proceedings of the International Society for Music Information Retrieval Conference.

[Gemmeke17] Gemmeke, Jort F et al. (2017). “Audio Set: An ontology and human-labeled dataset for audio events”. In: Proceedings of the Acoustics, Speech and Signal Processing International Conference.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement N° 688382