Deliverable D4.2

First prototype tool for the automatic semantic description of music samples

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	28 April 2017 (M15)
Submission date	28 April 2017 (M15)
Report availability	Public (X), Confidential ( )
Deliverable type	Report ( ), Demonstrator (X), Other ( )
Task leader	MTG-UPF
Authors	Dmitry Bogdanov, 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 Description of the annotation tool 4

1.1 Goals 4

1.2 Implementation 7

1.3 Usage instructions 7

2 Conclusion and future improvements 9

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

Executive Summary

As part of the Audio Commons Ecosystem, a number of tools are provided for the automatic analysis of audio content without the need for human intervention. These tools are designed for extracting i) musical audio properties for music pieces and music samples, and ii) non-musical audio properties for any kind of sounds. Work-in-progress versions of these tools have been released in parallel for the first prototype in the Audio Commons Ecosystem.

The current deliverable addresses the automatic extraction of musical audio properties for sound samples and demonstrates the first prototype of the tool for extracting such properties. The tool consists of an Essentia audio extractor which extracts a number of audio properties and a Python script which post-processes some of these properties and renames it to the commonly agreed descriptor names of the Audio Commons Ontology. The whole ensemble is provided as a Docker container which makes it really easy to be installed and executed on any platform.

The musical properties included in this audio extractor have been selected according to a draft sound schema for the Audio Commons Ontology (which has not yet been included in the ontology itself). The current version of the annotation tool does not include all specified descriptors but only a few of them. Future versions of the tool will include more descriptors. Other potential improvements include the unification of the current tool with the ones demonstrated in deliverables D4.3 and D5.3 (i.e. the tool for the annotation of musical properties for music pieces and the tool for the annotation of non-musical properties).

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

1 Description of the annotation tool

1.1 Goals

The Audio Commons Ecosystem will provide tools for the automatic annotation of different kinds of audio content. These tools are developed and evaluated in in different tasks across work packages 4 and 5. In particular, we will focus on the analysis of:

- Musical properties for music samples
- Musical properties for music pieces
- Non-musical audio properties (timbral models)

The differentiation between music samples and music pieces has been done to tailor some descriptors to the particular context in which they're used. We define music pieces as audio recordings typically corresponding to *complete* songs, while we define music samples as individual or simpler music elements such as single notes, percussive hits, chords, melodies or loops. Note that loops (or short music fragments) can also be classified under music pieces in case they're complex enough.

As part of the development of the Audio Commons Ontology, a sound schema is being developed in which we specify, among others, a number of musical and non-musical properties which could be included in the automatic annotation tools. The following table shows the musical properties that are considered for both music pieces and music samples, and highlighted the ones which are relevant to the present deliverable. For details on non-musical audio properties we refer the reader to the deliverables D5.1 Hierarchical ontology of timbral semantic descriptors and D5.2 First prototype of timbral characterisation tools for semantically annotating non-musical content.

Name	Description	Relevant for...
`ac:duration`	Duration of audio	Music pieces + Music samples
`ac:format`	Audio format (e.g. wav, mp3)	Music pieces + Music samples
`ac:lossless`	Whether audio file is in lossless codec (1 or 0)	Music pieces + Music samples
`ac:codec`	Codec used for encoding the audio (e.g. pcm_s16le)	Music pieces + Music samples
`ac:filesize`	Size of the file	Music pieces + Music samples
`ac:bitrate`	Number of bits per second	Music pieces + Music samples
`ac:bitdepth`	Number of bits per sample	Music pieces + Music samples

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

ac:samplerate	Number of samples per second	Music pieces + Music samples
ac:channels	Number of channels	Music pieces + Music samples
ac:audio_md5	MD5 checksum of raw undecoded audio payload. It can be used as a unique identifier of audio content.	Music pieces + Music samples
ac:genres	Music genre(s) of the musical content	Music pieces
ac:instruments	Musical instruments present in the recording	Music pieces + Music samples
ac:mood	Mood conveyed by the musical content	Music pieces
ac:tonality	Key and scale (e.g. A minor)	Music pieces + Music samples
ac:tempo	Tempo in BPM of the audio signal	Music pieces + Music samples
ac:chord	Played chord (e.g. G#m)	Music samples
ac:note	Played note name (e.g. C4)	Music samples
ac:midi_note	Played note midi number (e.g. 60)	Music samples
ac:loudness	Loudness value	Music pieces + Music samples
ac:dynamic_range	Dynamic range of audio recording	Music pieces + Music samples
ac:temporal_centroid	Temporal centroid	Music samples
ac:attack	Attack length	Music samples
ac:decay	Decay length	Music samples
ac:sustain	Sustain amount	Music samples
ac:release	Release length	Music samples
ac:log_attack_time	Logarithm of the time it takes to reach maximum amplitude of audio signal (good for perceptual attack)	Music samples

Highlighted in green are the descriptors which have already been implemented in this first version of the annotation tool. Highlighted in red are the musical properties which are not yet present in the tool.

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

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

1.2 Implementation

The annotation tool in particular can be found in the following public Github repository: https://github.com/AudioCommons/ac-audio-extractor. It consists of a "dockerized"[1] Python script which uses the Essentia Python bindings to compute audio descriptors and then reformats the output to match the property names defined in the Audio Commons sound schema. Essentia (http://essentia.upf.edu) is an audio analysis library developed and maintained at the Music Technology Group in Universitat Pompeu Fabra, coordinators of the Audio Commons Project. More information about the Essentia framework is provided in deliverable D4.3 First prototype tool for the automatic semantic description of music pieces.

The tool provides a command line utility to analyze a single audio file and write the output analysis into a JSON file (see next section for instructions). It does not manage the analysis of collections of audio files. This is expected to be done by another script that calls the command line utility and is run in the host machine. That other script is out of the scope of this deliverable.

1.3 Usage instructions

In order to use the analysis tool, Docker will need to be installed in the host system. If docker is installed the steps are quite simple:

1. Clone repository https://github.com/AudioCommons/ac-audio-extractor.
2. `cd` to repository folder and run `docker build -t audiocommons/ac-audio-extractor .`
3. Place the audio files that need to be analyzed in the same repository folder and run `docker run -it --rm -v \`pwd\`:/essentia audiocommons/ac-audio-extractor -i filename.wav -o analysis_output.json`

This will analyze the file `filename.wav` and place the results in `analysis_output.json`. In order to analyze files from other directories, one can either create a symbolic link in the same directory as the repository, or mount a different directory in the Docker container by replacing `` `pwd` `` with the path to that directory.

Running the above commands will produce an output like the following:


```
{
        "ac:bitrate": 705600,
        "ac:channels": 1,
        "ac:duration": 7.86884355545,
        "ac:dynamic_range": 2.23466777802,
        "ac:filesize": 694110,
        "ac:log_attack_time": -0.523798882961,
        "ac:lossless": 1,
        "ac:loudness": -12.3562965393,
        "ac:midi_note": 34,
        "ac:samplerate": 44100,
        "ac:tempo": 120,
        "ac:temporal_centroid": 0.487449944019,
        "ac:audio_md5": "5310f7f32bd87e460b9280322ef3fcbb",
```


[1] Docker (http://docker.com) is a multi-platform software that allows the deployment of software into containers. Containers work like lightweight virtual machines which include all the dependencies needed for the software to run. Therefore, "dockerized" programs are easy to run in different platforms as the Docker layer handles dependencies and other configuration parameters.

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


```
        "ac:codec": "pcm_s16le",
        "ac:note": "A#1",
        "ac:tonality": "A minor”
}
```


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

2 Conclusion and future improvements

In this demonstrator deliverable we described the first version of an automatic tool for annotating musical properties of music samples. We present the tool as a command line utility that can be given an input audio file and extract relevant musical properties according to a work-in-progress Audio Commons sound schema. The tool can run on different platforms as it is provided in a Docker container.

This tool is nevertheless expected to be continuously updated during the length of the Audio Commons project. In particular, deliverables D4.7 and D4.12 will describe improved versions of the tool, and deliverables D4.4 and D4.10 will provide evaluation results of its implemented algorithms. A list of improvements that are expected to be included in future iterations include:

- **Adding more descriptors from the sound schema**: some of the descriptors defined in the sound schema have not yet been included in the annotation tool (the ones highlighted in red in the tables above). Those will be implemented and added to the tool.
- **Improving quality of current descriptors**: some of the current descriptors are generated by basic implementations that could be improved. A good example is the pitch detection algorithm used for the "note" audio property. Future versions of the tool will feature intelligent segmentation of the pitch contour to improve pitch estimation.
- **Confidence measures**: following research that we've been carrying out[2] as part of Task 4.2 (RTD on automatic methods for the semantic annotation of music samples), future versions of the tool will incorporate confidence measures for some of the musical properties included in the audio analysis.
- **JSON-LD output**: also similarly to the evolution that the Audio Commons Mediator is expected to follow (see deliverable D2.5 Service integration technologies), future versions of this tool will be able to export to the JSON-LD format to enhance compatibility with semantic web technologies. Such output will be built by integrating the Audio Commons Ontology with the annotation tool.
- **Integration with other tools described in D4.3 and D5.2**: as mentioned above, the tool presented in this deliverable only covers a part of the automatic tools that are expected to be developed during the Audio Commons project. In future iterations we expect to fuse the developments of the three tools in a single "dockerized" command line tool that can be easily executed on any platform. The code presented in this deliverable can be used as the starting point for providing such combined tool.

[2] See paper published at ISMIR conference: Font, F., Serra, X. (2016). Tempo Estimation for Music Loops and a Simple Confidence Measure, published in "Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)".

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