D2.4 – High-level music content analysis framework v2

November 27th, 2019

Author/s: Rémi Mignot (IRCAM), Thomas Lidy (Musimap)

Contributor/s: Frédéric Cornu (IRCAM), Hadrien Foroughmand (IRCAM), Alvaro Thomas (Musimap)

Deliverable Lead Beneficiary: IRCAM

This project has been co-funded by the HORIZON 2020 Programme of the European Union. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

Deliverable number or supporting document title	D2.4 – High-Level Music Content Analysis Framework v2
Type	Prototype
Dissemination level	Public
Publication date	2019-11-30
Author(s)	Rémi Mignot (IRCAM), Thomas Lidy (Musimap), Frédéric Cornu (IRCAM), Hadrien Foroughmand (IRCAM), Alvaro Thomas (Musimap)
Contributor(s)	FuturePulse Consortium
Reviewer(s)	Christos Koutlis (CERTH), Daniel Johansson (SYB)
Keywords	Music Analysis, Content Analysis, Audio Analysis, Prediction from Audio, Tag prediction, Tag annotation, Deep Learning
Website	www.futurepulse.eu

CHANGE LOG

Version	Date	Description of change	Responsible
V0.1	05/11/2019	Copy of the content of D2.5.	Rémi Mignot (IRCAM)
V0.2	05/11/2019	Update of the SaaS of IRCAM	Frédéric Cornu (IRCAM)
V0.3	18/11/2019	Contributions: general + for Musimap	Thomas Lidy (Musimap)
V0.4	20/11/2019	Update of IRCAM analyses: introduction and performances.	Rémi Mignot (IRCAM)
V0.5	21/11/2019	Contributions Musimap	Thomas Lidy (Musimap)
V0.6	25/11/2019 27/11/2019	Update of Musimap API endpoints Corrections/completions after Review	Alvaro Thomas (Musimap), Thomas Lidy (Musimap)
V0.7	26/11/2019	Corrections after Review	Frédéric Cornu (IRCAM), Rémi Mignot (IRCAM)
V0.8	27/11/2019	Finalisation	Rémi Mignot (IRCAM)

Table of Contents

Funded by the European Commission

1 Executive Summary

Deliverable D2.4 “High-level music content analysis framework v2” is a prototype implementation of the music analysis and indexing framework available to the FuturePulse Platform and Applications. The present document is an accompanying report describing the prototype implementation.

Deliverable D2.4 is related to the task T2.4 “High-level music content analysis and indexing” involving IRCAM and Musimap. It is the update of D2.2 “High-level music content analysis framework v1”, which presented the state of T2.4 at the end of the first year of the project. The present deliverable D2.4 details the current status of both the results of audio content analyses and API specifications.

In the following sections we describe the implementations made by IRCAM (section 3) and by Musimap (section 4), and the integration in the FuturePulse platform (section 5).

Funded by the
European Commission

2 Introduction and Relation to other WPs/Tasks

WP2 of the FuturePulse project is concerned with collecting, analyzing and indexing data. Four sources of data have been identified and correspond to the four sub-tasks of WP2:

- data derived from broadcasting (T2.1)
- data derived from online music platform (T2.2)
- data derived from social media and open data (T2.3)
- data derived from the audio tracks themselves. These data are obtained using “Music Content Analysis” and denoted by “High-Level Music Content” data (HLMC) in the following (T2.4)

These four sources of information are then used jointly to perform the tasks of WP3 on predictive analysis and recommendations. The latter are used to allow the development of the three pilots (record label, live music and music platform) in WP5.

In deliverables D2.1 and D2.3, the data from broadcasting (T2.1), online music platform (T2.2) social media and open data (T2.3) has been specified. Those are data accessible through a Web API.

The first objective of T2.4 has been to allow the access of HLMC data through a Web API. This API provides the specific HLMC data needed and specified by the partners in charge of the three pilots. The second objective of T2.4 has been therefore to perform the specific research and development to allow the automatic estimation of these HLMC data from the audio.

In terms of organization, the requirements have been split in two sets and assigned to the two partners of T2.4 (IRCAM and Musimap). This led to the development of two Web APIs. We describe those APIs (and their development) in the following sections as well as their integration in the subsequent section.

Funded by the European Commission

3 IRCAM High-Level Music Content Analysis Framework

3.1 Introduction

The IRCAM API is constructed around a core C++ technology named **IrcamMusicDescription (IMD)**. The IRCAM API allows uploading an audio track (POST method) and getting their corresponding "high-level music content" (HLMC) information computed using IMD: BPM, Fade-in/Fade-out, duration, key, mode, tags of Electronic-Genre (as defined by BN during the first year), tags of Genre (as defined by SYB) and Vocal/Instrumental/Gender tags (as described later in Table 1). This information is then stored in a database for later retrieval using ISRC, and the uploaded audio file is removed.

Each user is assigned to a group. Users can only access information from tracks uploaded by users of the same group, and each user is assigned to a group by the FuturePulse platform. In the case several users from the same group upload the same track, computation and storage of HLMC will only be performed once. If users are from different groups, the computation and storage of HLMC will be performed again. This is because each group is allowed to have a specific configuration of HLMC.

This dynamic system allows processing the audio tracks from the three partners (SYB, PGM and SONAR) in the same way, therefore providing a uniform description to all audio; while preserving confidentiality (only data from a group are accessible to users from that group). Moreover increasing the audio collections of each partner is straightforward with this dynamic system.

What is of critical importance to mention is that in case the audio cannot be uploaded (for copyright reasons) to the IRCAM API, the core C++ technology can also be deployed and run directly on the partner's site ensuring the computation of the same "high-level music content" information.

3.2 Requirements Overview

The IRCAM framework for music content analysis (analysis of the audio content of a music file) is named IrcamMusicDescription (IMD). IMD is a C++ software available for Linux, MacOS and Windows.

In the FuturePulse project we have adapted the functionalities of IMD to fulfill the requirements of the use case partners of the project: SYB, PGM, SONAR (or BN in the past). The use case requirements assigned to IRCAM are given in Table 1. It should be noted that no requirement was formulated by PGM related to music content analysis.

Funded by the European Commission

Table 1 Use Case Requirements related to Music Content Analysis.

Requirement Code	Name	Description
BMP_REQ#4	Genre	Provides the genre of the track following SYB/PGM proposed taxonomy. *This one contains 1 level of 28 terms (African, Ambient, Bass, … Soul, Techno)*
LM_REQ#1	Genre of electronic music	Provides the sub-genre of the track following the BN proposed electronic music taxonomy of the first year. *This one contains 1 level of 22 terms (Ambient, Breakbeat, ...).*
BMP_REQ#12	BPM	Provides the number of Beats Per Minute (range from 10 bpm to 360bpm) of the track. It is supposed that the tempo does not change over the duration of the track. If it changes, it provides the median tempo value.
BMP_REQ#13	Fade in/fade out	Provides the duration in seconds of the Fade-in (if existing) and Fade-out (if existing).
BMP_REQ#14	Major/minor mode	Provides the mode (major or minor) of the musical key used in the track. It is supposed that the key does not change over the track duration. If it changes, it provides the most used key.
BMP_REQ#10	Vocal gender/ instrumental	Provides the gender (male/female) of the main singer(s) in the track. If there is no singer, it returns the value “instrumental”.

3.3 Adapting IMD for FuturePulse needs

Among the six requirements mentioned above, “BPM” and “major/minor mode” were already supported by IMD prior to the start of the FuturePulse project [Peeters2006] [Peeters2007]. Nevertheless, during the second year some improvements have been done.

For the “Fade In and Fade out” durations estimation, a method was designed for the project and integrated in IMD at the end of the first year. Another approach has been also developed during the second year of the project.

To address the remaining three requirements, new developments took place specifically for the project using the IrcamClassification [Peeters2015] machine learning framework.

Funded by the European Commission

3.3.1 Datasets

The datasets provided by SYB and BN to train the machine-learning models are described in Table 2.

Table 2 Datasets provided by SYB and BN to train machine-learning models

Requirement Code	Dataset name	Description	Distribution
BMP_REQ#4	SYB-Genre	Genre of a track	Around 4000 tracks annotated (multi-label) into SYB genres
BMP_REQ#10	SYB-Vocal	Vocal gender/ instrumental of a track	Around 4000 tracks annotated (single-label) into male, female, instrumental
LM_REQ#1	BN-Electronic genre	Genre of electronic music of a track	Around 6000 tracks annotated (single-label) into BN electronic music genre

3.3.2 Performance levels at M27

In the following we provide the performance levels obtained at M27 for the three tag estimation problems, the tempo estimation in BPM and the key/mode recognition.

3.3.2.1 SYB Genre of a track

Since a track can belong to multiple SYB genre-tags simultaneously (the tags are not mutually exclusive), the problem is solved as a multi-label classification.

At M27, using the above mentioned training set, the following results have been obtained: mean Recall: 71.83%, mean Specificity: 90.17%.

These results have been obtained using a 5-cross-validation with artist filtering. The "artist filtering" consists in gathering all the tracks of an artist into the same fold/group, in order to evaluate the ability to classify unknown artists. The "Recall" measures the number of times (in percent) a track tagged as A is correctly recognized as A. The "Specificity" measures the number of times (in percent) a track tagged as not-A is correctly recognized as not-A. Both measures are essential to characterize the performance of a system in the multi-label case.

Table 3 should be read as follows:

- Only 1% of the tracks are tagged as "african" (column "Tagged").
- For this 1% of tracks, 55.8% of them are correctly tagged by the system as "african" (column "Recall").
- For the remaining 99% of track, 98.8% of them are correctly tagged as "non-african" (column "Specificity").
- In total 98.4% of the tracks are correctly tagged as being African (when they are) or not African (when they are not) (column "Accuracy"). Note however, that since the dataset is highly unbalanced toward the negative classes, a high

value of Accuracy could be obtained even with a system that always output the negative class.

It should be noted that while the SYB/PGM taxonomy contains 60 genre labels; the dataset provided by SYB only provides annotation for 28 of them. Therefore Table 3 only represents the results for these 28 genre classes.

Table 3 Performances for the automatic recognition of SYB Genres (multi-label classification)

Labels	Tagged (%)	Recall (%)	Specificity (%)	Accuracy (%)
african	1	55,8	98,8	98,4
ambient	2,1	80,3	98,8	98,6
bass	0,8	25	98,9	98,2
blues	1,4	43,6	98,1	97,4
christian	2,9	75,6	91,4	90,9
classical	3,2	92,2	98,9	98,7
country	7,1	91,5	89,6	89,7
dance/edm	10,8	88,5	86,4	86,6
dancehall/reggaeton	2,4	90,4	96	95,9
disco	2,4	69,2	94,7	94,1
experimental	1,8	34,7	97,1	96
folk	4,4	62,9	91,4	90,2
funk	3,3	73,6	92,4	91,7
hiphop	5,2	86,6	94,9	94,5
house	8	87	90,5	90,2
indie	14,7	65,8	78,3	76,5
jazz	7,2	77,3	94,6	93,5
lounge	5,6	66,9	90,1	88,8
mariachi	0,3	50	100	99,8
pop	34,9	79,6	65	70,3
r&b	4,8	69,4	88,2	87,3
reggae/dub	0,7	38,7	99,7	99,2
rock	20,7	84,3	84,2	84,3
salsa	0,3	69,2	99,9	99,8
samba/bossa	1,2	66,7	98,2	97,8
singer-songwriter	6	78,7	87,1	86,6
soul	7,2	78,2	87,9	87,2
techno	1,1	64,2	97,4	96,9
mean	**5,77**	**69,50**	**92,45**	**92,11**

Funded by the European Commission

3.3.2.2 SYB-Vocal

Since a track can belong to only one SYB vocal-tag (tags are mutually exclusive), the problem is solved as a single-label classification.

Using the above training set, the following results have been obtained: mean Recall of 84.5%. For this 3-classes problem, a random classifier would lead to a 33% mean Recall. These results have been obtained using a 5-cross-validation with artist filtering.

Figure 1 presents the confusion between tags (to be read as 86% of the track tagged as “instrumental” are recognized as “instrumental”, 8% as “male”, 6% as “female”).

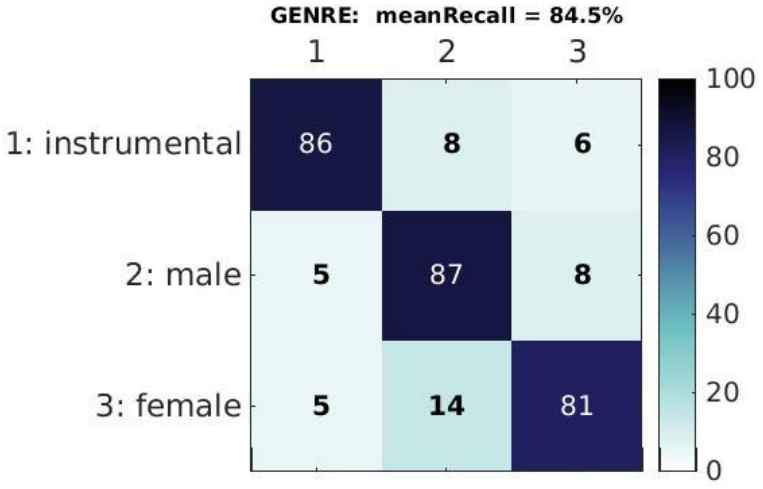

Figure 1 Confusion Matrix for the Instrumental/Gender automatic classification

3.3.2.3 BN-Electronic Genre

Since a track can belong to only one BN-electronic-genre-tag (tags are mutually exclusive), the problem is solved as a single-label classification.

Using the above mentioned training set, the following results have been obtained: mean Recall of 59.1%. For this 22-class problem, a random classifier would lead to a 4.5% mean-Recall. These results have been obtained using a 5-cross-validation with artist filtering.

Figure 2 presents the confusion between tags (to be read as 81% of the track tagged as “Trance” are recognized as “Trance”, 6% as “Techno”, 8% as “House”…).

Funded by the European Commission

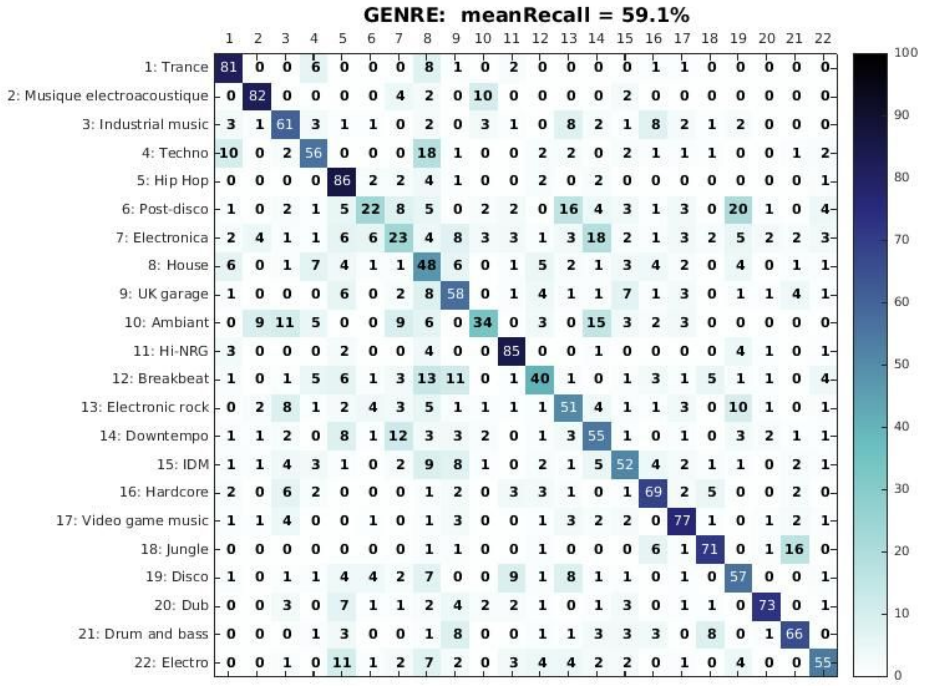

Figure 2 Confusion Matrix for the BN-Electronic-Genre automatic classification

3.3.2.4 Tempo in BPM

During the second year of the project, the initial method of [Peeters, 2006] has been improved by developing an approach that aims to reduce the number of octave errors. An error of octave is due to a confusion, for example: are the beats associated to crotchet notes or to minims notes, twice longer ? Then, two definitions of the tempo can exist with a factor 2. Even if the choice of humans can be consensual for a given recording, most algorithms can make a confusion.

Based on public datasets annotated in BPM for research purposes, the method of [Peeters, 2006] and its improvement have been evaluated and compared. As seen in Table 4, the exact accuracy has been improved by 2.2 points, and accepting an error margin of 4%, by 5 points. Nevertheless, as expected no improvement has been obtained when the octave errors are acceptable. This method has been implemented in IMD.

The method of [Foroughmand, 2019] is a new method based on Convolutional Neural Networks, and has been developed as part of the project. It has provided a significant improvement by almost 10 points compared to [Peteers, 2006] with a margin error of 4%. This method will be implemented in IMD by the end of the project.

Table 4 Results of Tempo estimation

Measure	[Peeters, 2006]	[Peeters, 2006] with improvements	[Foroughmand, 2019]
Exact accuracy	36.7 %	**38.9 %**	36.8 %
Accuracy ±4%	64.6 %	69.6 %	**74.4 %**
Accuracy ±4% + octave errors	88.9 %	88.9 %	**92 %**

3.3.2.5 Key and Mode

The first implemented method was based on [Peeters, 2007]. During the second year, another method [Papadopoulos, 2011] has been selected for developments, with some new improvements.

Using datasets annotated in global key and mode, publicly available for research purposes, the methods [Peeters, 2007] and [Papadopulos, 2011] with improvements, have been evaluated and compared.

Table 5 presents the results of two different scores: “exact key / mode” measures the percentage of good predictions by assuming that both the key and the mode names must be correctly recognised; and “mode recognition” only considers the mode name (e.g. predicting C Major instead of G Major is counted as correct). The new implementation has provided significant improvements for the first measure, by more than 5 points, and small improvement for the second measure, almost 1 point.

Table 5 Results of Key / Mode recognition

Measure	[Peeters, 2007]	[Papadopoulos, 2011] with improvements
Exact key / mode	48.6 %	**54.9 %**
Mode recognition	72.1 %	**73.2 %**

Funded by the European Commission

3.4 Bringing the technology into a SaaS

Within the FuturePulse project the C++ IMD software has been made accessible as a service through the development of a RESTful Web API described below.

The web service is written in Python using Flask, Werkzeug and is connected to a MongoDB database. All data (ISRC, group, and music content analysis data) is stored in the MongoDB database.

The aim of the web service is to provide access to the High-Level Music Content (HLMC) data of a music track.

For this, the web service has two main methods:

1. POST: to ingest new data in the MongoDB database,
2. GET: to retrieve the HLMC data given an ISRC.

3.4.1 User/group

Each user is assigned to a group on the FuturePulse platform. Ingestion of new audio track and Retrieval of HLMC data are specific to each group. A user can only ingest or retrieve data from its own group.

All methods require an authorization token specified in the Authorization header. A validated token supplies the user's group.

3.4.2 POST method AUDIO

The POST method allows ingesting a new audio track with its corresponding ISRC.

When a track is uploaded (POST), the corresponding HLMC data and ISRC are stored in a placeholder only accessible to the users of the same group.

When a track is uploaded, the web service first checks (using md5 identification) if this track has been previously uploaded by a user of the same group.

- If the file has not been already uploaded: the HLMC data of the ingested audio track are computed using IMD and stored in a MongoDB database with the corresponding ISRC. This ISRC becomes the access key for later retrieving the associated HLMC data.
- If the file has already been uploaded: the HLMC data are not re-computed and only the potentially new ISRC information is stored. It is therefore possible to assign different ISRC to the same audio track as it happens in reality.

3.4.2.1 Example

The following uploads the audio track `/user/futurepulse/audiofile/USUM71604430.mp3` and stores the HLMC data with the ISRC access key USUM71604430.

```
curl -i -H 'Authorization: Bearer xxx.yyy.zzz' -X POST -F
"file=@/user/futurepulse/audiofile/USUM71604430.mp3"
url:5000/api/v2.0/track/isrc/USUM71604430
```


Funded by the European Commission

3.4.2.2 Output for new tracks

When ingesting an audio track for the first time, the output is the following

```
{   "success": {     "message": "new track added"   } }
```

3.4.2.3 Output for existing tracks

When ingesting an audio track which already exists in the user group, the output is the following

```
{   "success": {     "message": "new ISRC added for already existing track"   } }
```

3.4.2.4 Error output

When ingesting an audio track with a wrong audio format, the output is the following

```
{   "error": {     "code": "803",      "message": "invalid audiofile format"  } }
```

3.4.3 Notes on ISRC

The ISRC becomes the access key for later retrieval of HLMC data.

- Using the above mechanism it is possible that the same track (same md5) receives different ISRC.
- Using the above mechanism it is possible that different tracks (different md5) receive the same ISRC.
- Using the above mechanism, it is also possible that the same audio track (same md5) receives different ISRC in different user groups.
- It is left to the users to check the correctness of the ISRC assignment.

3.4.4 GET method ISRC

The first GET method allows retrieving the HLMC data corresponding to a given ISRC

3.4.4.1 Example

The following retrieves the HLMC data for the ISRC key USUM71604430.

```
curl -i -H ‘Authorization: Bearer xxx.yyy.zzz’
url:5000/api/v2.0/track/isrc/USUM71604430
```

3.4.4.2 Normal output

In case of success, the output has the following structure

```
{
  "success": {
    "data": [
      {
        "content_bpm": "140.9980628497",
```


Funded by the European Commission

```
"content_cuetype_fade-in": "0.0000000000",
"content_cuetype_fade-out": "29957.7777778",
"content_duration": "30013.3333333",
"content_key": "Bb",
"content_mode": "maj",
"content_tag": {
  "BNElectronicGenre": [
    {
      "confidence": "0.2550955210",
      "value": "Downtempo"
    },
    {
      "confidence": "0.1315994377",
      "value": "Hip Hop"
    },
    {
      "confidence": "0.1245708897",
      "value": "Disco"
    },
    {
      "confidence": "0.0811220070",
      "value": "House"
    },

  ...
  ],
  "SYBGenre": [
    {
      "confidence": "0.8840784556",
      "value": "singer-songwriter"
    },
    {
      "confidence": "0.7983233606",
      "value": "folk"
    },
    {
      "confidence": "0.6043557271",
      "value": "lounge"
    },
    {
      "confidence": "0.5536250746",
      "value": "indie"
    },

  ...
  ],
  "VocalGender": [
    {
      "confidence": "0.7147898493",
      "value": "male"
    },
    {
      "confidence": "0.1558780334",
      "value": "female"
    },
    {
      "confidence": "0.1293321173",
      "value": "instrumental"
    }
  ]
},
"filename": "blues.00030.wav",
```


Funded by the European Commission

```
          "groupID": "bmat",
          "isrc": [
            "11111",
            "USUM71604430"
          ],
          "md5": "e9e4ea0f9eb911dce0260617b4d06122"
        }
      ]
    }
}
```


3.4.4.3 Error output

In case no HLMC data are available for this ISRC (this is the case when the track/ISRC has never been ingested by a user of the same group), the output is the following


```
HTTP/1.0 404 NOT FOUND Content-Type: application/json Content-Length: 58 Server:
Werkzeug/0.14.1 Python/2.7.14 Date: Fri, 22 Jun 2018 16:50:24 GMT  {   "error": {
"message": "data does not exist"   } }
```


3.4.5 GET method ALLTRACKS

The second GET method allows to retrieve all the HLMC data corresponding to all the tracks owned by the group the user belongs to.

3.4.5.1 Example

The following retrieve the HLMC data for all the tracks assigned to the group the user belongs to. A subset can be selected using the page and num_per_page query parameters

```
curl -i -H ‘Authorization: Bearer’
url:5000/api/v2.0/alltracks?page=2&num_per_page=20
```

3.4.5.2 Normal output

In case of success, the output is similar as above given that data is a list of results


```
{
  "success": {
    "data": [

      {...},

      {...},

      {...},

      {...},

      {...},
    ]
  }
```


Funded by the European Commission

```
}
```

3.4.6 Error messages

Code	Context	Description
400		Invalid pagination syntax
403		Access forbidden (user has insufficient access privilege)
404		Not found (URL does not exist, ISRC resource has not yet been created)
803	Track creation	invalid audio file format

3.5 Possible formats for input audio data

The following audio formats can be used for the input audio file

Format	Description
wav	Waveform Audio File Format
mp3	formally MPEG-1 Audio Layer III or MPEG-2 Audio Layer III
flac	Free Lossless Audio Codec

3.6 Description of output data

Feature name	Description
content_bpm	Provide the average tempo of the track in beats per minute
content_cuetype_fadein	Provide the end time of the fade in in msec
content_cuetype_fadeout	Provide the start time of the fade out in msec
content_duration	Provide the time duration of the track in msec
content_key	Provide the average key of the track (C, Db, … Bb, B)
content_mode	Provide the average mode of the track (maj, min)
content_tag	Within each of the three tag families (BNElectronicGenre, SYBGenre, VocalGender) the tags are ranked by decreasing value confidence
--- confidence	Confidence/likelihood (between 0 and 1) of the tag for this track. Note for single-label problems, the confidences of the tags of a same family sum to 1. For multi-label problems, this is not the case.
--- value	Value of the tag

Funded by the European Commission

3.7 Tag dictionaries

SYBGenre		
african	experimental	r&b
ambient	folk	reggae/dub
bass	funk	rock
blues	hiphop	salsa
christian	house	samba/bossa
classical	indie	singer-songwriter
country	jazz	soul
dance/edm	lounge	techno
dancehall/reggaeton	mariachi	
disco	pop	

VocalGender		
female	instrumental	male

BNElectronicGenre		
Ambient	Electronica	Musique electroacoustique
Breakbeat	Hardcore	Post-disco
Disco	Hi-NRG	Techno
Downtempo	Hip Hop	Trance
Drum and bass	House	UK garage
Dub	IDM	Video game music
Electro	Industrial music	
Electronic rock	Jungle	

Funded by the European Commission

4 Musimap High-Level Music Content Analysis Framework

4.1 Requirements Overview

The use case requirements assigned to MUSIMAP are presented in Table 6.

Table 6 Use Case Requirements related to music content analysis.

Requirement Code	Name	Description
BMP_REQ#4	Genre	Provides the genre of the track following by multi-genre detection of Musimap's 67 genres taxonomy.
BMP_REQ#11	Moods	Provides the moods associated to a specific track.
LM_REQ#17	Emotional analysis of artists	Provides moods associated to an artist. Derived either from the artist's portrait or accumulated from analysis of the tracks of an artist
BMP_REQ#6	Energy level	Provides information about perceived energy in a track.
BMP_REQ#10	Vocal gender/ instrumental	Provides the gender (Male/female) of the main singer(s) in the track. If there is no singer, it returns the value "instrumental". In all 3 cases a percentage of confidence / intensity is provided.
BMP_REQ#12	BPM	Provides the number of Beats Per Minute of the track. It is supposed that the tempo does not change over the duration of the track. If it changes, it provides the median tempo value.

4.2 Audio Data Sets

Two data sets have been provided to cover the different use cases in FuturePulse, with respect to audio analysis:

- SYB data set: ~ 37,000 audio files
- PGM data set: ~ 51,000 audio files

These two sets of audio files have been analyzed with the criteria listed in Table 6 and described in the following sub-sections. Each set of audio files is accompanied by a metadata CSV file. A sample of such a CSV file with column description is shown in Table 7. Note that for the PGM data set no Youtube URLS have been provided.

Table 7 Sample of CSV file containing track related information.

Track	Artist(s)	Album	ISRC	Spotify URI	YouTube URL
Track's name also known as title	Track's artist(s)	Album name	International Standard Recording Code: a unique code used to identify a track	A unique Spotify URI containing the Spotify track's ID	A YouTube URL containing the YouTube track's id

Funded by the European Commission

4.3 Audio Analysis Attributes

The following subsections describe the attributes (tags) Musimap analyzes and predicts from a track.

4.3.1 Vocal gender / instrumental

Musimap has built a deep-learning based system that employs the latest methods in detecting whether a track is instrumental or vocal. The system is based on a Convolutional Neural Network (CNN) architecture. The input to the network are Mel Spectrograms extracted from different segments of the audio file under analysis. The network outputs the probability of the presence of a vocal gender or instrumentalness.

This system has been trained and tested on Musimap's corpus with hundreds of thousands of tracks, which has been manually labelled by musicologists, and provides a high accuracy. Vocal gender (male/female) recognition has been added in year 2 of FuturePulse. It detects the presence of male or female singing voice or both by outputting a percentage level for both genders.

4.3.2 Moods

Musimap has a database of annotated tracks with up to 256 moods (words related to the human emotions). The mood taxonomy is structured in:

- 6 main mood families
- 18 sub moods (3 per mood family)
- 35 predicted complex moods
- 232 detailed complex moods
- 5 rhythmic moods (lyrical, flowing, staccato, stillness, chaotic)

Example:

- Rhythmic moods: staccato, chaotic
- Main Mood: Out (Wood)
    - sub-mood: 'manliness'
        - complex mood: 'wild'

An overview of the full mood taxonomy is available from: https://moods.musimap.net.

Currently, Musimap's mood tagging system can annotate automatically the 6 mood families, 18 sub-moods, and 35 complex moods plus 5 rhythmic moods, from the audio signal, i.e. in total up to 64 moods per song.

Each mood tag comes with an importance level in %. This is an integer value between 0 and 100 that describes the strength of each mood present in the track. Such an importance value has been used in the manual annotation of Musimap's training corpus and it is also available for the results of mood annotation, i.e. each mood assigned to a song will also have a value in % describing the strength of this mood in the song.

Musimap strives to expand the selection and number of automatically detected moods in the next iteration of the mood model (in Year 3 of FuturePulse).

Funded by the European Commission

4.3.3 Emotional analysis of artists

Musimap is working on aggregating the moods of an artist's tracks to provide a mood profile for the artist. This assumes the availability of the audio for the tracks of the artist.

4.3.4 Genres

Similarly as for moods, Musimap owns a corpus that has been annotated with up to 400 genres. The full genre taxonomy is available via: https://genres.musimap.com/

A Deep Learning based neural network (Convolutional Neural Network) has been trained to provide automatic genre annotation for 67 genres (a selection of the full taxonomy that is deemed the genres most important for music industry). These 67 genres are listed in Table 8. Musimap strives to expand the selection and number of automatically detected genres in the next iteration of the genre model (in Year 3 of FuturePulse).

Table 8 Musimap Genre Taxonomy

Musimap Genres		
Alternative Rock	Electric Blues	Pop Rock
Ambient New Age	Electro Body Music	Prelude
Big Band	Electro Pop	Punk
Blues	Eurodance	R&B - Nu Soul
Blues Rock	Experimental Electronic Music	Reggae
Bollywood	Folk	Reggaeton
Bop	Funk	Rock
Breakbeat	Gospel	Rock N Roll
Chanson Francaise	Hard Rock	Rock New Wave
Chill Out Lounge	Hip Hop	Rural Blues
Church Music	House	Salsa - Afro Cuban
Classical	IDM - Electronica	Samba - Bossa Nova - MPB
Concerto	Illbient Dark Ambient	Soul
Country	Indie Rock	Soul Jazz
Country Rock	Industrial	Speech
Dance Pop	Jazz	Swing
Dancehall	Latin Pop - Latin Dance	Symphony
Disco	Metal	Tech House
Drum & Bass	New Wave	Techno
Dub	Noise Music	Trip Hop
Dubstep	Nu Metal	Vocal Jazz

Funded by the European Commission

Easy Listening	Opera
EDM	Pop

4.3.5 BPM

Musimap also provides beats per minute detection from the audio signal. A tempo estimation is performed on the song. It is supposed that the tempo does not change over the duration of the track. If it changes, the output will be the median tempo value.

4.3.6 Energy level

The Energy level of a track represents the energy value throughout the audio signal of a track.

It is analyzed through one of the low-level audio analysis libraries used at Musimap. We employ different audio analysis tools and algorithms which analyze the spectrum of the audio file and perform an onset and beat detection to derive e.g. tempo, beats per minute, loudness etc. Energy level is also a result of this low-level spectral analysis. The result value is directly provided on track level via API.

4.4 API Usage

In this section we describe the usage of Musimap's API to retrieve the Musimap audio analysis results via the accessible endpoints made available for the FuturePulse platform.

4.4.1 Audio Analysis Results: Get Track Analysis Data

MusiMap's audio analysis creates metadata through analysis a track by the aforementioned parameters and providing tags for the tracks including genres, moods, voice family and musical attributes such as, energy, key, mode and BPM.
The following endpoint allows to retrieve the predicted metadata from an analyzed track:

Resource URL:

```
GET https://api.musimap.io/musimotion
```

Parameters:

`reference` (required) : string // identifier of the source track, e.g. ISRC

`version` (optional): string // analyzer version in use (currently 1.2)

Funded by the European Commission

Request sample:

```
curl -X GET \ 'https://api.musimap.io/musimotion?reference=USIR10210955' \
```

In this example the sample song called "A Thousand Miles" has the USIR10210955 identifier (ISRC).

Response example:


```
{
    "status": 200,
    "timestamp": 1574763232791,
    "data": {
        "mood": [
            {
                "score": 63.722044,
                "item": {
                    "id": "f1aea1a6-ca6a-4b20-b0f6-5a8ea50f6cb6",
                    "name": "Friendly"
                },
                "value": 63.722044
            },
            {
                "score": 54.54163,
                "item": {
                    "id": "2154f72e-2004-45b4-8360-9200634c92e8",
                    "name": "Sentimental"
                },
                "value": 54.54163
            },
            {
                "score": 51.9672705,
                "item": {
                    "id": "a03a702b-0ec4-300a-d18c-1542c9023315",
                    "name": "Good vibrations"
                },
                "value": 51.9672705
            },
          [...]
        ],
        "acoustic_attribute": [
            {
                "score": 268.794189453125,
                "item": {
                    "id": "63588566-5b72-4d99-a685-1aade8dfe964",
                    "name": "Duration"
                },
                "value": 268.794189453125
            },
            {
                "score": 94.98788452148438,
                "item": {
                    "id": "582ba553-1b91-4425-88e4-d58432c9fee5",
                    "name": "Bpm"
                },
                "value": 94.98788452148438
            },
            {
                "score": 33.5345512869966,
```


Funded by the European Commission

```
                "item": {
                    "id": "93a6f555-450e-4d1f-b3b4-0631d71f7441",
                    "name": "Rhythm Presence"
                },
                "value": 33.5345512869966
            },
            {
                "score": 9.750043671380341,
                "item": {
                    "id": "06132ef1-2688-4032-9232-01a699193748",
                    "name": "Spectral Energy"
                },
                "value": 9.750043671380341
            },
            {
                "score": 2.0,
                "item": {
                    "id": "82fe3702-fca7-4493-8499-1f574d446acc",
                    "name": "Key"
                },
                "value_label": "B",
                "value": 2.0
            },
            {
                "score": 1.0,
                "item": {
                    "id": "2a246dc3-e21f-4806-a91e-e6d621436310",
                    "name": "Key Scale"
                },
                "value_label": "major",
                "value": 1.0
            }
        ],
        "genre": [
            {
                "score": 75.578505,
                "item": {
                    "id": "20152b88-901e-4c57-867f-1ebe2176be40",
                    "name": "Pop"
                },
                "value": 75.578505
            },
            {
                "score": 38.983792449999996,
                "item": {
                    "id": "e1e8d1a3-bc09-41bd-8108-954b839d7f94",
                    "name": "Folk"
                },
                "value": 38.983792449999996
            },
            {
                "score": 37.37945,
                "item": {
                    "id": "41cab628-887b-495d-86c9-29f783b70bef",
                    "name": "Country"
                },
            },
          [...]
        ],
        "rhythmic_mood": [
            {
                "score": 50.3501,
                "item": {
                    "id": "eba5e088-c286-d5e7-80a5-a1eae337b4d9",
```


Funded by the European Commission

```
                    "name": "Staccato"
                },
                "value": 50.3501
            },
            {
                "score": 31.435000000000002,
                "item": {
                    "id": "6316bc1f-d794-4bc9-6a8c-fbe3868664f6",
                    "name": "Flowing"
                },
                "value": 31.435000000000002
            },
            {
                "score": 13.380851999999999,
                "item": {
                    "id": "095a8c30-3f6e-2a7d-ff6c-e7866676ea98",
                    "name": "Lyrical"
                },
                "value": 13.380851999999999
            }
        ],
        "voices_family": [
            {
                "score": 91.80663,
                "item": {
                    "id": "2f0c46c8-03c1-c9c0-cb49-fca8bdb0b84f",
                    "name": "Female Vocal"
                },
                "value": 91.80663
            },
            {
                "score": 0.0,
                "item": {
                    "id": "0b897061-b94f-477f-9d72-06491f14a8de",
                    "name": "Instrumental (no vocal)"
                },
                "value": 0.0
            },
            {
                "score": 0.0,
                "item": {
                    "id": "2c967e7e-d9ee-817b-a332-ec8e5984ae50",
                    "name": "Male Vocal"
                },
                "value": 0.0
            }
        ],
    }
}
```


When the track is successfully analyzed, an HTTP 200 status response is delivered together with the following information:

- acoustic attributes:
    - Key: key of the song (as a note from A to G)
    - BPM: tempo in beats per minute
    - Duration: in seconds

Funded by the European Commission

- Spectral Energy: energy contained in the signal, equal to the summation across all frequency components of the signal spectral density
- Rhythm Presence: how danceable the track is (a percentage value between 0 & 100, higher values meaning more danceable)

- moods: moods in the song (bitter, atmospheric, energetic, etc)
- genres: song category
- rhythmic moods: chaotic, flowing, lyrical, staccato, stillness
- voices family: instrumental (0 or 100%), or female/male or both with a % value of intensity

4.4.2 Get Moods Taxonomy

The following endpoint returns the Musimap moods taxonomy:

Resource URL:

```
https://api.musimap.io/lexicology/
```

Request example:

```
curl -X GET https://api.musimap.io/lexicology/?category=mood&product=musimotion:v1.2
```

Response example::

```
{
    "status": 200,
    "timestamp": 1574847436555,
    "data": [

        {
            "id": "9be268c7-674b-4bd3-a2a3-a8883d347aa7",
            "category_name": "mood",
            "name": "Aggressive",
            "description": "Parent mood : Manliness
\nPrecise meaning : Assertive, quarrelsome or belligerent
\nMusical example :  System Of A Down - B.Y.O.B.           "
        },


        {
            "id": "46deb385-ef44-4bdf-95a3-d51730898ac1",
            "category_name": "mood",
            "name": "Atmospheric",
            "description": "Parent mood : Happiness
\nPrecise meaning : Aerial, ethereal, light
\nMusical example :  Harold Budd & Brian Eno - The Pearl"
        },
```


Funded by the European Commission

```
        {
            "id": "78fbe83e-7b97-4ac0-a772-c8e2a66ab58d",
            "category_name": "mood",
            "name": "Bitter",
            "description": "Parent mood : Coldness
\nPrecise meaning : Resentful and disenchanted
\nMusical example : Nick Cave & Bad Seeds - The moon Is In The gutter           "
        },

        {
            "id": "894fd98a-eaaa-48b8-aedd-334c6273972a",
            "category_name": "mood",
            "name": "Cerebral",
            "description": "Parent mood : Intellect
\nPrecise meaning : Inner sophistication, intricate thinking
\nMusical example : Keith Jarrett - Koln concert"
        },

        {
            "id": "c56b7569-62b0-9bda-a116-40dc13465d16",
            "category_name": "mood",
            "name": "Coldness",
            "description": "Meaning all moods related to : BITTERNESS & CYNICISM
\nMoments : Moral isolation / Nasty thoughts
\nMoods providing : Depression / Sarcasm
\nMusical example : Cold wave"
        },

        {
            "id": "bfcd0f69-0f0f-43b1-8916-1eaa33bb8c98",
            "category_name": "mood",
            "name": "Cool",
            "description": "Parent mood : Nourishment
\nPrecise meaning : Relaxed and easy going
\nMusical example : Bob Marley - Redemption song"
        },

        {
            "id": "28e18fda-89a3-4670-8c4c-ba9df201e2e3",
            "category_name": "mood",
            "name": "Depressed",
            "description": "Parent mood : Withdrawal
\nPrecise meaning : Closure facing life
\nMusical example : The Cure - Faith"
        },
        {
            "id": "128f1a5d-9589-4be0-8211-45f038de8f40",
            "category_name": "mood",
            "name": "Determined",
            "description": "Parent mood : Warrior
\nPrecise meaning : Brave and active
\nMusical example : Bruce Springsteen - Born in the USA"
        },

        {
            "id": "1a6d56ca-3757-4eeb-8cc2-ade01108c65a",
            "category_name": "mood",
            "name": "Dreaming",
            "description": "Parent mood :  Imagination
\nPrecise meaning : Thoughtful  creative floating
\nMusical example :  This Mortal Coil - Song to the siren"
```


Funded by the European Commission

```
        },


        {
            "id": "1e66c428-f242-7510-c474-303c768870f4",
            "category_name": "mood",
            "name": "Dynamism",
            "description": "Meaning all moods related to : ENERGY / ENTHUSIASM
\nMoments : Catharsis / Letting go
\nMoods providing : Exultation / Rhythm drive
\nMusical example : Uptempo latin dance"
        },

        {
            "id": "27241d72-b61c-4f22-b390-01ac87bc1489",
            "category_name": "mood",
            "name": "Energetic",
            "description": "Parent mood : Dynamism
\nPrecise meaning : Alive and kicking
\nMusical example : Mark Ronson & Bruno Mars - Uptown Funk"
        },

        {
            "id": "9b58478c-75c0-502b-a8d1-02eb242e8e38",
            "category_name": "mood",
            "name": "Extroversion",
            "description": "Meaning all moods related to : EXAGERATION / IMPETUOUSITY
\nMoments : Pretentiousness / Self confidence
\nMoods providing : Tacky attitude / Flashy
\nMusical example : Glam rock / Gangsta rap"
        },


        {
            "id": "d9675b29-b61c-44a7-9a51-443f533a1f35",
            "category_name": "mood",
            "name": "Free",
            "description": "Parent mood : Happiness
\nPrecise meaning : Instinctive way of acting out of any conventions     \nMusical
example :  Lambert & Hendricks & Bavan - Cousin Mary"
        },

        {
            "id": "f1aea1a6-ca6a-4b20-b0f6-5a8ea50f6cb6",
            "category_name": "mood",
            "name": "Friendly",
            "description": "Parent moods :  Good vibrations
\nPrecise meaning : Reassuring, confident and benevolent
\nMusical example : Jack Johnson - Better together"
        },
        {
            "id": "ddc0d458-e7ee-4ff8-9ebc-8996cee5faaf",
            "category_name": "mood",
            "name": "Funny",
            "description": "Parent mood : Temperament
\nPrecise meaning : Joyful smiles
\nMusical example : AronChupa - I m an Albatraoz              "
        },
        {
            "id": "bd3ac812-726f-473c-a29b-0c2234435d19",
            "category_name": "mood",
            "name": "Glamourous",
```


Funded by the European Commission

```
            "description": "Parent mood : Love\nPrecise meaning : alluring and
fascinating in a showy way\nMusical example : Nina Simone - My baby don t care"
        },
        {
            "id": "a03a702b-0ec4-300a-d18c-1542c9023315",
            "category_name": "mood",
            "name": "Good vibrations",
            "description": "Meaning all moods related to : SHARING / POSITIVE
COMMUNICATION                                                    \nMoments :
Generosity / Amicability
\nMoods providing : Good vibes
\nMusical example : Surf music"
        },
        {
            "id": "f05f72ad-e056-cea2-3bd2-e109efa0c02b",
            "category_name": "mood",
            "name": "Happiness",
            "description": "Meaning all moods related to : POSITIVITY / FREEDOM
\nMoments : Festivity / Easy living
\nMoods providing : Unworried feelings / Innocent amusement              \nMusical
example : Teen pop"
        },
        {
            "id": "3bf7fd61-d4a7-4cbc-8286-3fe4d17a3adf",
            "category_name": "mood",
            "name": "Happy",
            "description": "Parent mood : Happiness
\nPrecise meaning : Innocent, Good side of life
\nMusical example : Pharrell Williams - Happy"
        },
         {
            "id": "499de113-eedc-4863-8588-582a248be545",
            "category_name": "mood",
            "name": "Heroic",
            "description": "Parent mood :   Warrior
\nPrecise meaning : Courageous but desperate
\nMusical example :  Richard Wagner - Ride of the Valkyries"
        },

        {
            "id": "5900660d-85a0-4ce2-a129-4b5b3bbab08b",
            "category_name": "mood",
            "name": "Humorous",
            "description": "Parent mood : Temperament
\nPrecise meaning : Displaying and creating smiles and laughs
\nMusical example : Louis Prima - Just a gigolo"
        },
         {
            "id": "8a74b98a-887c-a9bc-ecdf-a495765820be",
            "category_name": "mood",
            "name": "Imagination",
            "description": "Meaning all moods related to : MAGIC & ENCHANTMENT
\nMoments : Fantasy / Dreaming
\nMoods providing : Floating / Graceful
\nMusical example : Erik Satie"
        },
        {
            "id": "35ce3ef9-68ca-42b7-a22d-afe312ff8d09",
            "category_name": "mood",
            "name": "Impulsive",
            "description": "Parent mood : Extroversion
\nPrecise meaning : acting on sudden desires rather than careful thoughts
\nMusical example : James Brown - Sex machine"
```


Funded by the European Commission

```
        },

        {
            "id": "54485e53-65fa-4d74-adf4-4ac03089c949",
            "category_name": "mood",
            "name": "Inhibited",
            "description": "Parent mood : Withdrawal
\nPrecise meaning : Shyness
\nMusical example : Joao Gilberto - Desafinado"
        },
        {
            "id": "e47037a9-cd12-4936-8c63-4c3795adb120",
            "category_name": "mood",
            "name": "Innocent",
            "description": "Parent mood : Happiness
\nPrecise meaning : Naïve and positive way of facing life with minimum fuss
\nMusical example : Bobby McFerrin - Don t worry be happy"
        },
        {
            "id": "148c18f3-b7f5-4719-bf30-5364d41d7e8c",
            "category_name": "mood",
            "name": "Inspired",
            "description": "Parent mood : Self control
\nPrecise meaning : Dedicated to gods
\nMusical example : Jeff Buckley - Hallelujah"
        },
        {
            "id": "aec666ef-1869-a54e-18f0-a8f87fb1fec8",
            "category_name": "mood",
            "name": "Intellect",
            "description": "Meaning all moods related to : SOPHISTICATION &
COMPLEXITY                                                    \nMoments : Analytical
reflection / Inside exploration
\nMoods providing : Introspection
\nMusical example : Electro-acoustic music / Free jazz"
        },
         {
            "id": "2cc908b7-d8dd-42b7-9f26-f102130b05c5",
            "category_name": "mood",
            "name": "Lively",
            "description": "Parent mood : Happiness
\nPrecise meaning : Full of life and vivacious
\nMusical example : Antonio Vivaldi - Spring (Allegro pastorale)
"
        },
        {
            "id": "6755a82e-689a-e825-ee02-5551736cdb4d",
            "category_name": "mood",
            "name": "Love",
            "description": "Meaning all moods related to : SEDUCTION & SENSUALITY
\nMoments : Desire / Openneness
\nMoods providing :  Lustful feeling / Affection
\nMusical example : Quiet storm soul music"
        },
            "id": "92e96fa3-74b5-75b2-27c1-96b7d22d4bbb",
            "category_name": "mood",
            "name": "Manliness",
            "description": "Meaning all moods related to :  ROUGH MASCULINITY /
WILDNESS                                                        \nMoments : Rebellion
/ Hostility
\nMoods providing : Threats / Violence
\nMusical example : Extreme metal"
        },
```


Funded by the
European Commission

```
        {
            "id": "7bde440f-a510-416b-90c2-70a67c98ae1b",
            "category_name": "mood",
            "name": "Meditative",
            "description": "Parent mood : Spirituality
\nPrecise meaning : Zen
\nMusical example : Lama Gyurme - Song of Awakening"
        },
        {
            "id": "7da6c979-d799-4d54-8195-a0a407a664fc",
            "category_name": "mood",
            "name": "Melodramatic",
            "description": "Parent mood : Temperament
\nPrecise meaning : Overacted climax
\nMusical example : Klaus Nomi - Cold song  "
        },
        {
            "id": "1beeca8a-0184-f79b-6924-5188daf241bd",
            "category_name": "mood",
            "name": "Nourishment",
            "description": "Meaning all moods related to : SHELTER & COMFORT
\nMoments : Warm and cool atmospheres
\nMoods providing : Feeling safe / Comfy
\nMusical example : Bluesy vocal jazz"
        },

        {
            "id": "f5e21d03-2e33-4f7c-b758-000c7daaf97c",
            "category_name": "mood",
            "name": "Organic",
            "description": "Parent mood : Roots
\nPrecise meaning : Feeling in the guts
\nMusical example : Muddy Waters - Mannish boy"
        },

        {
            "id": "384dd193-47a9-462c-8ec8-56f33d91e661",
            "category_name": "mood",
            "name": "Passionate",
            "description": "Parent mood : Love
\nPrecise meaning : Passionate love affair    ,Intense , scorching
\nMusical example : Moody Blues - Night in white satin"
        },
        {
            "id": "86c8394b-0aa9-4b5f-a6f5-83fb1e76e06e",
            "category_name": "mood",
            "name": "Performing",
            "description": "Parent mood : Playfulness
\nPrecise meaning : Giving one self (in a communicative pleasure)         \nMusical
example : Deep Purple - Smoke Under Water (Live in Japan)"
        },
        {
            "id": "9ffc5764-3ee0-1c96-a840-4ee4fd0f974f",
            "category_name": "mood",
            "name": "Playfulness",
            "description": "Meaning all moods related to : PROVOCATION & PERFORMANCE
\nMoments : Ready to face life
\nMoods providing : Competitivity / Self affirmation
\nMusical example : Pop punk"
        },
        {
            "id": "8a6f03d1-9b5c-40d1-a2a0-7a91ed3e30b6",
            "category_name": "mood",
```


Funded by the European Commission

```
            "name": "Powerful",
            "description": "Parent mood : Warrior
\nPrecise meaning : Omnipotent and vigorous
\nMusical example :  Survivor - Eye of the tiger"
        },
        {
            "id": "69802afb-40bc-4f9a-acd7-2a9ba0ae5bd1",
            "category_name": "mood",
            "name": "Rebellious",
            "description": "Parent mood : Manliness
\nPrecise meaning : Challenging social rules
\nMusical example : Sex Pistols - Anarchy in the UK"
        },
        {
            "id": "62f1b60e-95b0-4164-8be1-5bc66c886b44",
            "category_name": "mood",
            "name": "Refined",
            "description": "Parent mood : Sensibility
\nPrecise meaning : Subtle, soft or fragile beauty
\nMusical example :  Frantz Schubert - Piano sonata n°6"
        },
        {
            "id": "a8e784e7-60cb-4a9a-bc67-ed2b99a8c8dc",
            "category_name": "mood",
            "name": "Revitalizing",
            "description": "Parent mood : Dynamism
\nPrecise meaning : Positive driving energy
\nMusical example : Michael Jackson - Bad"
        },
        {
            "id": "2294612b-1233-4f34-b5b6-bdcea772b1cc",
            "category_name": "mood",
            "name": "Roots",
            "description": "Meaning all moods related to : ANCHORAGE & ORGANIC
\nMoments : Back to basics / Tribalism
\nMoods providing : Prosaicness / Core feelings
\nMusical example : African ritual & Trance music"
        },
        {
            "id": "d87e763c-a97e-4ac8-9723-ad6dbc8dd8ee",
            "category_name": "mood",
            "name": "Sad",
            "description": "Parent mood : Sensibility
\nPrecise meaning : Feeling sorrow or unhappy
\nMusical example :  Lana Del Rey - Pretty When You Cry"
        },
        {
            "id": "38161f57-8d19-7675-03a6-51505bc8261a",
            "category_name": "mood",
            "name": "Self control",
            "description": "Meaning all moods related to : MYSTICISM & RELIGION
\nMoments : in communication with the above, praying
\nMoods providing : Ecstasy / Inspiration
\nMusical example : Gregorian chants"
        },
        {
            "id": "6d5ae713-fa4a-2067-6cc8-b296d29ef57b",
            "category_name": "mood",
            "name": "Sensibility",
            "description": "Meaning all moods related to : ROMANTICISM & FEMINITY
\nMoments : Privacy / Platonic love souvenirs
\nMoods providing : Fragility / Melancholia
\nMusical example : Fado"
```


Funded by the European Commission

```
        },
        {
            "id": "f74046c8-7d3b-41e5-bdc4-b040c7983064",
            "category_name": "mood",
            "name": "Sensual",
            "description": "Parent mood : Love
\nPrecise meaning : Physical desire and attractiveness
\nMusical example : Donna Summer - Love to love you baby"
        },
        {
            "id": "2154f72e-2004-45b4-8360-9200634c92e8",
            "category_name": "mood",
            "name": "Sentimental",
            "description": "Parent mood : Sensibility
\nPrecise meaning : Romantic and melancholic
\nMusical example : Chet Baker - Almost blue"
        },
        {
            "id": "0158d0de-e706-4d42-a727-ee006188c5e1",
            "category_name": "mood",
            "name": "Soothing",
            "description": "Parent mood : Nourishment
\nPrecise meaning : Feeling self-care and sheltered (and cured)
\nMusical example : Stan Getz & Joao Gilberto - The girl from Ipanema"
        },

        {
            "id": "633210e7-d8c2-9e3c-07a6-03968d0298ac",
            "category_name": "mood",
            "name": "Spirituality",
            "description": "Meaning all moods related to : MEDITATION
\nMoments : Inner peace / Ready to contemplate
\nMoods providing : Balance / Harmony
\nMusical example : New age / Relaxation music"
        },
        {
            "id": "c63ce1d1-6178-7984-8629-6c70b180270a",
            "category_name": "mood",
            "name": "Temperament",
            "description": "Meaning all moods related to : SPONTANEOUSITY / WITTINESS
\nMoments : Kitsch / Melodrama
\nMoods providing : Freshness / Vulnerability
\nMusical example : Operetta (Opera Buffa)"
        },
        {
            "id": "2cc67cfa-8c2d-44f1-b045-7d62608cd88a",
            "category_name": "mood",
            "name": "Warm-hearted",
            "description": "Parent mood : Good Vibrations
\nPrecise meaning : Positive and warmful communication
\nMusical example : Louis Armstrong - What a wonderful world"
        },
        {
            "id": "ce8cfa4a-2fd0-6408-7497-c141256c7489",
            "category_name": "mood",
            "name": "Warrior",
            "description": "Meaning all moods related to : CONQUEST & PROUDNESS
\nMoments : Heroism / Vigor
\nMoods providing : Solidity / Determination
\nMusical example : Star Wars theme"
        },
        {
            "id": "f695778e-6475-f312-350c-66114b1bdec8",
```

```
            "category_name": "mood",
            "name": "Withdrawal",
            "description": "Meaning all moods related to : SOLITUDE & INTROVERSION
\nMoments : Doubt / Inhibition
\nMoods providing : Closure / Organisation
\nMusical example : Drone or lowercase music"
        }
    ],
    "offset": 0,
    "limit": -1,
    "size": 197
}
```


Having a look at one of the delivered moods, for example:


```
        {
            "id": "3bf7fd61-d4a7-4cbc-8286-3fe4d17a3adf",
            "category_name": "mood",
            "name": "Happy",
            "description": "Parent mood : Happiness
\nPrecise meaning : Innocent, Good side of life
\nMusical example : Pharrell Williams - Happy"
        },
```


For each mood the following points are provided:

- "id": uid assigned to each mode
- "category_name": mood (as this is the category searched through the endpoint)
- "name": name of the delivered mood
- "description": further details on the given mood with three fields
    - Parent mood: main mood where the children moods are related to
    - Precise meaning: mood definition
    - Musical example: a song where this mood can be found

Funded by the European Commission

5 Conclusions and integration on the FuturePulse platform

Deliverable D2.4 “High-level music content analysis framework v2” is the implementation of the music analysis and indexing framework available to the FuturePulse platform. This document is an accompanying report describing the respective prototype implementations of IRCAM and Musimap, and their integration.

Some HLMC data are covered by both web APIs. It is the case for: estimated value of the tempo (BMP_REQ#12), recognition of key/mode (BMP_REQ#14), tags for the vocal gender/instrumental classification (BMP_REQ#10), and tags of Genres (BMP_REQ#4, with two different genre taxonomies). To avoid redundant information, it is planned that the FuturePulse platform integrates these different information, either by filtering or by merging.

Based on informal evaluations conducted by musicologists, the following decisions have been made: The estimated tempo and the key/mode recognition of IRCAM are selected to be provided on the FuturePulse platform, ignoring the values returned by Musimap. On the contrary, for the vocal gender/instrumental classification, the displayed tags are obtained through the Musimap API, not from IRCAM. The latter choice has been done due to better performances of the Musimap analysis, and also for its ability to recognise mixed songs (with both male and female voices).

For the genre classification, because IRCAM and Musimap use different taxonomies, a merging procedure will be developed on the FuturePulse platform at ATC. The general idea is the following: For genres which appear in both taxonomies, the used probability of a recording to belong to a genre will be computed by averaging the corresponding probability estimates of IRCAM and Musimap. For genres which only appear in one taxonomy, the corresponding probability will appear without change.

Note that some genres of IRCAM include two or more genres of Musimap. For example "DanceHall/Reggaeton" of IRCAM includes "Dancehall" and "Reggaeton" that are separated at Musimap. In those cases, it has been decided to only keep the Musimap genres because they provide more details. Finally note that also the IRCAM genres "christian" and "indie" will be removed from the platform.

Table 9 presents the merged genre taxonomy that will appear on the FuturePulse platform, with for all genre the corresponding source. Note that the symbol “*AVG(Ircam, Musimap)*” means that the used value is the **average** of the probability estimates obtained from the IRCAM API and the Musimap API.

Funded by the European Commission

Table 9 Ircam + Musimap Genre Taxonomy and API source.

Ircam + Musimap MERGED	API Source:	Ircam + Musimap MERGED	API Source:
African	IRCAM	IDM - Electronica	Musimap
Alternative Rock	Musimap	Illbient Dark Ambient	Musimap
Ambient New Age	AVG(Ircam, Musimap)	Indie Rock	Musimap
Big Band	Musimap	Industrial	Musimap
Blues	Musimap	Jazz	AVG(Ircam, Musimap)
Blues Rock	Musimap	Latin Pop - Latin Dance	Musimap
Bollywood	Musimap	Mariachi	IRCAM
Bop	Musimap	Metal	Musimap
Breakbeat	Musimap	New Wave	Musimap
Chanson Francaise	Musimap	Noise Music	Musimap
Chill Out Lounge	AVG(Ircam, Musimap)	Nu Metal	Musimap
Church Music	Musimap	Opera	Musimap
Classical	AVG(Ircam, Musimap)	Pop	AVG(Ircam, Musimap)
Concerto	Musimap	Pop Rock	Musimap
Country	AVG(Ircam, Musimap)	Prelude	Musimap
Country Rock	Musimap	Punk	Musimap
Dance Pop	Musimap	R&B - Nu Soul	AVG(Ircam, Musimap)
Dancehall	Musimap	Reggae	Musimap
Disco	AVG(Ircam, Musimap)	Reggaeton	Musimap
Drum & Bass	Musimap	Rock	AVG(Ircam, Musimap)
Dub	Musimap	Rock N Roll	Musimap
Dubstep	Musimap	Rock New Wave	Musimap
Easy Listening	Musimap	Rural Blues	Musimap
EDM	AVG(Ircam “dance/edm”, Musimap “EDM”)	Samba - Bossa Nova - MPB	AVG(Ircam, Musimap)
Electric Blues	Musimap	Salsa - Afro Cuban	AVG(Ircam, Musimap)
Electro Body Music	Musimap	Singer-Songwriter	IRCAM
Electro Pop	Musimap	Soul	AVG(Ircam, Musimap)
Eurodance	Musimap	Soul Jazz	Musimap
Experimental Electronic Music	AVG(Ircam, Musimap)	Speech	Musimap
Folk	AVG(Ircam, Musimap)	Swing	Musimap
Funk	AVG(Ircam, Musimap)	Symphony	Musimap

Funded by the European Commission

Gospel	Musimap
Hard Rock	Musimap
Hip Hop	AVG(Ircam, Musimap)
House	AVG(Ircam, Musimap)

Tech House	Musimap
Techno	AVG(Ircam, Musimap)
Trip Hop	Musimap
Vocal Jazz	Musimap

Funded by the
European Commission

6 References

[Peeters, 2006] Geoffroy Peeters. Chroma-based estimation of musical key from audio-signal analysis. In Proc. of ISMIR (International Society for Music Information Retrieval), pages 115–120, Victoria, BC, Canada, 2006.

[Peeters, 2007] Geoffroy Peeters. Template-based estimation of time-varying tempo. EURASIP Journal on Applied Signal Processing, 2007(1):158–158, 2007. doi:10.1155/2007/67215.

[Peeters 2015] Geoffroy Peeters, Frederic Cornu, David Doukhan, Enrico Marchetto, Remi Mignot, Kevin Perros, and Lise Regnier. When audio features reach machine learning. In Proc. of ICML (International Conference on Machine Learning), volume Machine Learning for Music Discovery Workshop, Lille, France, July, 6–11 2015.

[Papadopoulos, 2011] Hélène Papadopoulos, and Geoffroy Peeters. Local key estimation from an audio signal relying on harmonic and metrical structures. IEEE Transactions on Audio, Speech, and Language Processing, 20(4), 1297-1312, 2011.

[Foroughmand, 2019] Hadrien Foroughmand, and Geoffroy Peeters, Deep-rhythm for tempo estimation and rhythm pattern recognition, ISMIR, Delft, Netherlands, Nov. 2019.