D2.2 – High-level music content analysis framework v1

July 11th, 2018

Author/s: Geoffroy Peeters (IRCAM), Hadrien Foroughmand (IRCAM)

Contributor/s: Maria Stella Tavella (Musimap), Thomas Lidy (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	D2.2 – High-Level Music Content Analysis Framework v1
Type	Prototype
Dissemination level	Public
Publication date	2018-08-31
Author(s)	Geoffroy Peeters (IRCAM), Hadrien Foroughmand (IRCAM), Maria Stella Tavella (Musimap), Thomas Lidy (Musimap)
Contributor(s)	FuturePulse Consortium
Reviewer(s)	Vasilis Papanikolaou (ATC) Akis Papadopoulos (CERTH)
Keywords	Music Content Analysis, Indexing Framework, Prototype Implementation
Website	www.futurepulse.eu

CHANGE LOG

Version	Date	Description of change	Responsible
V0.1	01/07/2018	First version with part of IRCAM contribution	Geoffroy Peeters
V0.2	03/07/2018	First version with part of Musimap contribution	M. Stella Tavella
V0.3	10/07/2018	Second version of Musimap contribution	Thomas Lidy
V0.4	11/07/2018	Second version of IRCAM contribution	Geoffroy Peeters
V0.6	18/07/2018	Modifications following CERTH review	Geoffroy Peeters
V.07	19/07/2018	Modifications following CERTH review (MMAP section)	Thomas Lidy

Funded by the European Commission

Table of Contents

Funded by the European Commission

Funded by the
European Commission

1 Executive Summary

Deliverable D2.2 “"High-level music content analysis framework v1” 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 D2.2 is related to the task T2.4 “High-level music content analysis and indexing” involving IRCAM and Musimap.

In the following we describe the prototype implementation made by IRCAM (section 3) and by Musimap (section 4).

Funded by the
European Commission

2 Introduction and Relation to other WPs/Tasks

2.1 Purpose and Scope

WP2 of the FuturePulse project is concerned with collecting, analysing 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 deliverable D2.1, 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 during the first year, was to also allow the access of HLMC data through a Web API. This API should provide the specific HLMC data needed and specified by the partners in charge of the three pilots. The second objective of T2.4 during the first year was therefore to perform the specific research and development to allow the automatic estimation of these HLMC data from the audio.

In terms of organisation, the requirements has 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.

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), 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.

Each user is assigned to a group. Users can only access information from tracks uploaded by users of the same group. 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 group, 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 BN) 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 fulfil the requirements of the use case partners of the project (SYB, PGM, BN). The use case requirements assigned to IRCAM are given in the Table 1. It should be noted that no requirement was formulated by PGM related to music content analysis.

Requirement Code	Name	Description
SYB14	Genre	Provides the genre of the track following SYB/PGM proposed taxonomy. *This one contains only 1 level of 60 terms (African, Alternative, Ambient, Americana, … Trash Metal, Tri-pHop)*
BN1	Genre of electronic music	Provides the genre and sub-genre of the track following the BN proposed electronic music taxonomy. *This one contains 2 levels.* *Level 1 (Family) contains 5 terms (Lowtempo, Breaks, Electronic, …).* *Level 2 (Sub-Genre) contains 226 terms (Ambient dub, Ambient Industrial, Dark Ambient, …)*
SYB12	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

Funded by the European Commission

		the duration of the track. If it changes, it provides the median tempo value.
SYB13	Fade in/fade out	Provides the duration in seconds of the Fade-in (if existing) and Fade-out (if existing).
SYB14	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.
SYB10	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".

Table 1 Use Case Requirements related to music content analysis.

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].

To address the remaining four requirements, new developments took place specifically for the project using the IrcamClassification [Peeters-2015] machine learning framework.

3.3.1 Datasets

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

Requirement Code	Dataset name	Description	Distribution
SYB-14	SYB-Genre	Genre of a track	Around 4000 tracks annotated (multi-label) into SYB genres
SYB-10	SYB-Vocal	Vocal gender/ instrumental of a track	Around 4000 tracks annotated (single-label) into male, female, instrumental
BN-1	BN-Electronic genre	Genre of electronic music of a track	Around 6000 tracks annotated (single-label) into BN electronic music genre

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

3.3.2 Performance levels at M12

In the following we provide the performance levels obtained at M12 for the three tag estimation problems.

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 M12, using the above mentioned training set, the following results have been obtained: meanRecall: 71.83%, meanSpecificity: 90.17%.

Funded by the European Commission

These results have been obtained using a 5-cross-validation with artist filtering. The "Recall" measures the number of times (in percent) a track tagged as A is correctly recognised as A. The "Specificity" measures the number of time (in percent) a track tagged as not-A is correctly recognised as not-A. Both measures are essential to characterise 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 track 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.

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

Funded by the European Commission

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**

Table 3 Performances obtained at M12 for the automatic recognition of SYB Genres (multi-label classification)

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.

At M12, 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 represents the confusion between tags (to be read as 86% of the track tagged as “instrumental” are recognised as “instrumental”, 8% as “male”, 6% as “female”).

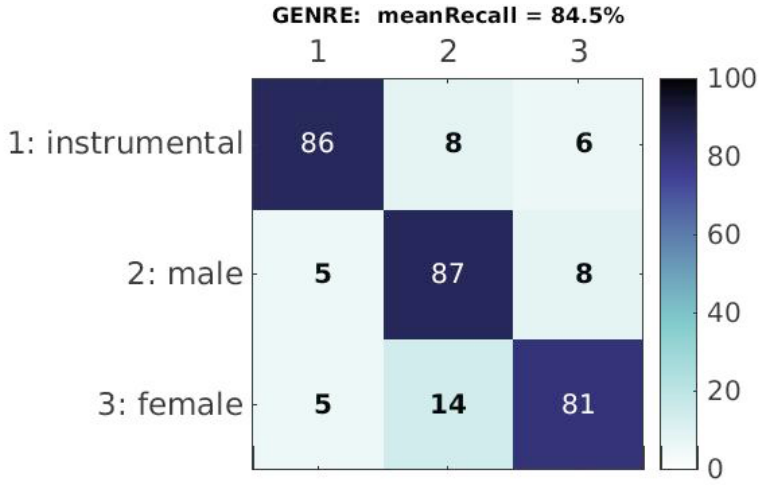

Figure 1 Confusion Matrix for the Instrumental/Gender automatic classification

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.

At M12, 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 represents the confusion between tags (to be read as 81% of the track tagged as “Trance” are recognised as “Trance”, 6% as “Techno”, 8% as “House”…).

Funded by the European Commission

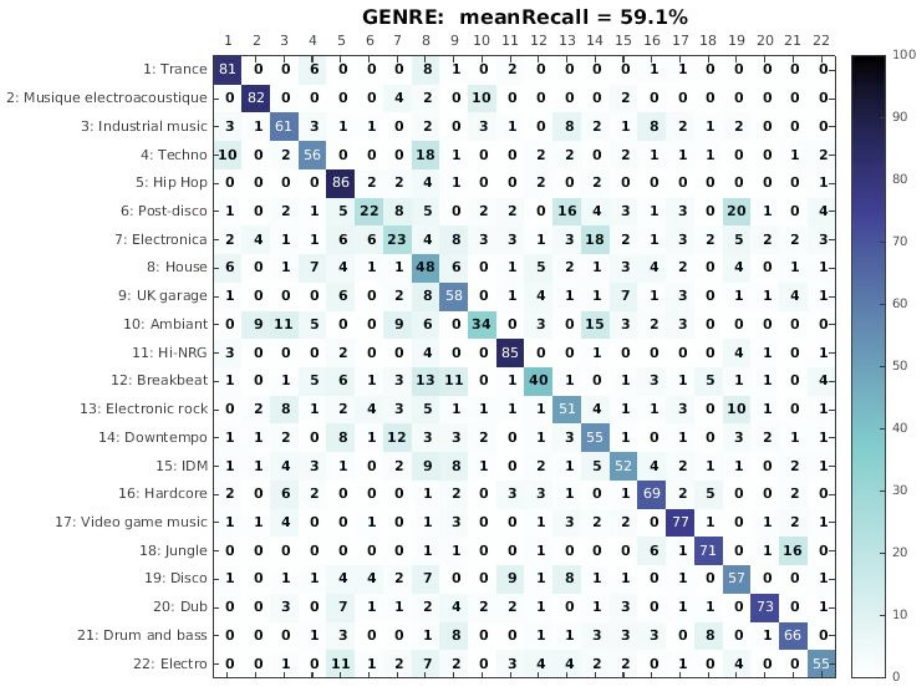

Figure 2 Confusion Matrix for the BN-Elctronic-Genre automatic classification

3.4 Bringing the technology into a SaaS

Within the FuturePulse project the C++ IMD software has been made accessible as a service trough 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 (user management and music content analysis data) are 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

The web service is password protected. It is accessible through the https (HyperText Transfer Protocol Secure) protocol. Only the hashes of the passwords are stored in the database.

Each user is assigned to a group. 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.

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.

Funded by the European Commission

- 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 -k -u $email:$password -X POST -F
"file=/user/futurepulse/audiofile/USUM71604430.mp3"
url:5000/api/v1.0/track/isrc/USUM71604430
```

3.4.2.2 Normal output

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

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

3.4.2.3 Error output

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

Funded by the European Commission

3.4.4.1 Example

The following retrieves the HLMC data for the ISRC key USUM71604430 of the group of the user $email.

```
curl -i -k -u $email:$password  url:5000/api/v1.0/track/isrc/USUM71604430
```

3.4.4.2 Normal output

In case of success, the output is the following


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


Funded by the European Commission

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

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 $email belongs to:

```
curl -i -k -u $email:$password  url:5000/api/v1.0/track
```

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

```
        {…},
        {…},
      ]
    }
}
```

3.4.5.3 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"   } }
```

3.4.6 Error messages

Code	Context	Description
401		Un-identified user
403		Access forbidden (user has un-sufficient access privilege)
404		Not found (URL does not exist, ISRC resource has not yet been created)
800	Group creation	group_ID already exist
801	Token creation	invalid group_ID
802	User creation	invalid token_ID
803	Track creation	invalid audio file format
804		User account has expired

3.5 User management

3.5.1 Creation of a user group

This method allows creating a new user group with a specific validity date and software configuration while it requires the administrator access privileges.

3.5.1.1 Examples

The following creates a new group named “`bmat`” with a validity up to the `2020-12-31` date and a software configuration named “`futurepulse201806`”.

```
curl -i -k -u $email_admin:$password_admin -X POST -F "validity=2020-12-31" -F
"configuration=futurepulse201806" url:5000/api/v1.0/admin/group/bmat
```

3.5.1.2 Normal output

```
{   "success": {     "message": "group_ID created"   } }
```


Funded by the European Commission

3.5.1.3 Error output

```
{   "error": {     "code": "800",      "message": "group_ID already exist"   }
}
```

3.5.2 Creation of an invitation token

To invite a new user it is necessary to create an invitation token. Each token is assigned to a group (hence a validity date and a configuration)

This method requires the administrator access privileges.

3.5.2.1 Example:

The following create an invitation token for the group "`bmat`".

```
curl -i -k -u $email_admin:$password_admin url:5000/api/v1.0/admin/token/bmat
```

3.5.2.2 Normal output

```
{   "success": {     "data": "38e7a00d15b94d44b20fe372d1741663"   } }
```

3.5.2.3 Error output

```
{   "error": {     "code": "801",      "message": "invalid group_ID"   } }
```

3.5.3 Creation of a user account

Given a token, a user can create an account by providing an email and a password. The access right of the user account are the one defined in the group.

3.5.3.1 Example

Using the provided invitation token (`38e7a00d15b94d44b20fe372d1741663`), a new user can be created by providing an $email and a $password.

```
curl -i -k -F "email=$email" -F "password=$password"
url:5000/api/v1.0/admin/user/38e7a00d15b94d44b20fe372d1741663
```

3.5.3.2 Normal output

```
{   "success": {     "message": "user created"   } }
```

3.5.3.3 Error output

```
{   "error": {     "code": "802",      "message": "invalid token_ID"   } }
```

3.5.3.4 Error output

```
{   "error": {     "code": "803",      "message": "email is already
used"   } }
```


Funded by the European Commission

3.6 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

3.7 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 (BNElectronic-Genre, 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

3.8 Tag dictionaries

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

Funded by the European Commission

Trance
UK garage
Video game music

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

VocalGender
female
instrumental
male

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 10.

Table 10. Use Case Requirements related to music content analysis.

Requirement Code	Name	Description
SYB10	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”.
SYB11	Moods	Provides the moods associated to a specific track.
BN 17+18	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
SYB6	Energy level	Provides information about perceived energy in a track.

4.1.1 Audio

The analysed audio files are the ones provided by FuturePulse partner SYB. The dataset consists of approximately 35,000 audio files, each of which contains several information collected in a .csv file. A sample of .csv file with column description is shown in Table 11:

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

Table 11, Sample of .csv file containing track related information.

4.1.2 Vocal gender/ instrumental

Musimap 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 vocal.

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.

Musimap will further improve it to detect also edge cases such as a background choir part in the middle of an otherwise instrumental song.

The addition of gender (male/female) recognition is planned for year 2 of FuturePulse.

Funded by the European Commission

4.1.3 Moods

Musimap provides a collection of 256 moods (words related to the human emotions) from the sociological lexicology. Moods are structured in 6 main mood families, 18 sub moods (3 per mood family) and 232 complex moods.

5 types of weighted rhythmic moods (e.g., lyrical, flowing, staccato, stillness, chaotic) complete the mood description of each track.

Example:

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

The full mood taxonomy is available from: https://moods.musimap.net and through the API (see below).

Currently Musimap's mood tagging system can annotate automatically the 6 mood families and 18 sub-moods, and a selection of 50 of the complex moods, from the audio signal.

As a result, an average of 50 moods and mood families describes each track.

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 track will also have a value in % describing the strength of this mood in the track.

4.1.4 Emotional analysis of artists

Musimap can aggregate 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.

For a number of well-known artists (so called "ambassadors" in different musical genres throughout music history), direct mood profiles of the artists are available, through manual annotation efforts. This will however not be available for new and upcoming artists, as here we rely on automated ingestion.

4.1.5 Energy level

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

It is analysed 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.

4.2 API Usage

In this section we will describe the common usage of Musimap's API and the accessible endpoints made available for the FuturePulse platform.

Funded by the European Commission

4.2.1 Instrumental/vocal

This endpoint allows retrieving information on whether a track is instrumental or vocal.

Resource URL:

```
https://api.musimap.net/tracks/search
```

Get information about whether a track is instrumental or vocal:

This is achieved by providing a flag to output the “details” of a track:

```
curl -H "Authorization: Bearer <encoded_access_token>” -XGET 'https://api.mu-
simap.net/tracks/search?limit=1&output=details'
```

Sample response: The Boolean output “instrumental” will be either true or false.


```
{“results":[
{"uid":"EA7C38BF-3449-5C62-E884-59CEDFCE761C",
"name":"Step It Pon the Rastaman Scene",
"duration":204,
"bpm":117,
"index":"01\/12",
"year":"06\/2003",
"best_of_jury":false,
"live":false,
"instrumental":false,
"bonus":false,
"hidden":false,
"hit_parade":false,
"unrecorded":false,
"score":100}]}
```


Another possibility is to filter the instrumental/vocal tracks by providing a Boolean value, in this example we filter by ‘instrumental=True’ and get 5 instrumental tracks (this will be any 5 instrumental tracks (5 specified via ‘limit’ parameter); to search for combination with other criteria, other API methods have to be combined with this one).

```
curl -H "Authorization: Bearer <encoded_access_token>” -XGET 'https://api.mu-
simap.net/tracks/search?limit=5&instrumental=True’
```

Sample response:


```
"results":[
{"uid":"DF222C5D-D9FA-40A8-5690-7EF1AD014A53","name":"January","score":0},
{"uid":"9C9F4664-124B-2A21-AA4F-EF8013C94C57","name":"Bezcitny","score":0},
{"uid":"2CEC9C6B-EC4D-FB3A-07DA-CD91218A51CC","name":"Pressure
Torture","score":0},
{"uid":"C30FE7DD-AB3D-AA9A-EB5A-A935D37C7688","name":"Lover Man","score":0},
```


Funded by the European Commission

```
{"uid":"4B88D963-6FCD-DBE8-DC65-4B1CE60F398D","name":"British IDM Preset Fanfare (The Hawaiian Hockey Song)","score":0}]}
```


4.2.2 Energy level

This endpoint serves to get the energy level of a track.

Resource URL:

```
https://api.musimap.net/tracks/search
```

Sample request:

```
curl -H "Authorization: Bearer <encoded_access_token>" -XGET 'https://api.musimap.net/tracks/search?limit=1&output=audioanalysis'
```

Sample response:


```
"results":[{"uid":"A8C437AD-7FCA-C8CD-DE04-D299FA98C2FD","name":"Rendezvous at Sunset","audioanalysis":{"key":
{"key":"G","scale":"minor","strength":"0.65127158165"},"derivative":"-0.000360064965207","danceability":"1.10073626041","centroid":"0.575814604759",
"duration":"212.559997559","complexity":"2.27092289925","loudness":"-15.0098733902","energy": "5581.37109375","intensity":"0"},"score":100}]
```


4.2.3 Get moods hierarchy

The call returns the moods hierarchy.

Resource URL:

```
https://api.musimap.net/moods/
```

Get Moods hierarchy

```
curl -H "Authorization: Bearer <encoded_access_token>" -XGET 'https://api.musimap.net/moods/'
```

In the curl replace sample values indicated by < > with your actual values.

Sample response:


```
{
    "results": [
        {
            "uid": "8E4F875D-7087-64ED-DECA-5295038D9A12",
```


Funded by the European Commission

```
            "name": "Above (Fire)",
            "children": [
                {
                    "uid": "38161F57-8D19-7675-03A6-51505BC8261A",
                    "name": "self control",
                    "children": [
                        {
                            "uid": "515746E4-4781-D94D-D962-57924B6438FF",
                            "name": "absolute",
                            "children": []
                        },
                        {
                            "uid": "10566D87-115D-CC00-3239-488AED4C2613",
                            "name": "vertical",
                            "children": []
                        },
                        {
                            "uid": "216A6949-6569-7369-7ECC-55523A3FF749",
                            "name": "flawless",
                            "children": []
                        },
                        {
                            "uid": "8CE7BC69-1374-D131-3284-B50986ED8462",
                            "name": "seeking",
                            "children": []
                        },
                        {
                            "uid": "57EDB459-31A5-727C-3702-369473AB1521",
                            "name": "inspired",
                            "children": []
                        }
                    ]
                },
                {
                    "uid": "633210E7-D8C2-9E3C-07A6-03968D0298AC",
                    "name": "spirituality",
                    "children": [
                        {
                            "uid": "3419CB61-7C05-DD3B-432E-98F2C463B403",
                            "name": "intuitive",
                            "children": []
                        },
…
```


Where:

- uid is the unique identifier code assigned, in this case, to each mood.
- name is the name of the mood.
- children are the sub-moods related to the main mood.

4.2.2. Get moods for a track

This method returns the moods associated with a track.

Resource URL:

```
https://api.musimap.net/tracks/
```

Get moods for a specific track. The track is identified by its unique uid:

Funded by the European Commission

```
curl -X GET \
 'https://api.musimap.net/tracks/search?reference=79F68A3C-1880-7843-0AC9-
31C18FC1FCB1&output=moods' \
 -H 'authorization: Bearer <encoded_access_token>” ' \
 -F reference=79F68A3C-1880-7843-0AC9-31C18FC1FCB1 \
 -F 'output=[moods]'
```

Sample response:


```
{
    "total": 1,
    "limit": 10,
    "offset": 0,
    "results": [
        {
            "uid": "79F68A3C-1880-7843-0AC9-31C18FC1FCB1",
            "name": "Demons",
            "moods": [
                {
                    "uid": "57EDB459-31A5-727C-3702-369473AB1521",
                    "name": "inspired",
                    "importance": 22
                },
                {
                    "uid": "8B3FF0C8-C5D1-56ED-CA65-49909BCA2D81",
                    "name": "essential",
                    "importance": 33
                },
                {
                    "uid": "55BEDE8D-6CC4-0C4A-F362-15A8560151C8",
                    "name": "organic",
                    "importance": 33
                },
                {
                    "uid": "1C1BC590-FD60-DDBA-D89E-558BE4CB6C68",
                    "name": "loose",
                    "importance": 55
                },
                {
                    "uid": "984A4E07-7EBC-717E-BF1E-64E48A5C5960",
                    "name": "letting go",
                    "importance": 37
                },
                {
                    "uid": "515746E4-4781-D94D-D962-57924B6438FF",
                    "name": "absolute",
                    "importance": 24
                },
                {
                    "uid": "B634F842-B62B-345A-F644-47B9BEE6D6A9",
                    "name": "weighty",
                    "importance": 21
                },
                {
                    "uid": "6B2F1E41-F8A9-DBAE-1910-19E99A0A1B55",
                    "name": "liberating",
                    "importance": 72
                },
                {
                    "uid": "38161F57-8D19-7675-03A6-51505BC8261A",
                    "name": "self control",
                    "importance": 9
```


Funded by the European Commission

```
        },
        {
            "uid": "446DDB2A-513B-0BDB-B2FD-6F25C401428A",
            "name": "thoughtful",
            "importance": 43
        },
        {
            "uid": "3A9A56D6-5138-0E95-3B40-4C1C795D1680",
            "name": "happy",
            "importance": 43
        },
        {
            "uid": "AFA2F319-53CC-EC1A-7F30-BF3B0F951AD8",
            "name": "driving",
            "importance": 33
        },
        {
            "uid": "2BE15410-02D8-7604-1972-88C229BE9261",
            "name": "sexual",
            "importance": 43
        },
        {
            "uid": "13F575EB-9BD7-BC37-61A8-4E80DED7F1AD",
            "name": "warm-hearted",
            "importance": 43
        },
        {
            "uid": "10936331-149D-BFEA-F8EA-70CFE8B2836D",
            "name": "positive",
            "importance": 43
        },
        {
            "uid": "B45AD70C-F647-D61D-7542-072AFDE3A526",
            "name": "flashy",
            "importance": 33
        },
        {
            "uid": "3212CFA7-0B20-4EC4-CD3D-08E291F6620A",
            "name": "serene",
            "importance": 47
        },
        {
            "uid": "458E616F-1B47-310F-755B-CA07EA19E8B6",
            "name": "proud",
            "importance": 27
        },
        {
            "uid": "0F009D11-53FD-52C7-FCF3-B12A853DA171",
            "name": "determined",
            "importance": 28
        },
        {
            "uid": "0599F73B-6A25-E9DA-4668-589A71739A73",
            "name": "pretentious",
            "importance": 28
        },
        {
            "uid": "7D4A6714-A918-43F7-D1F5-96A1963DFC00",
            "name": "in your face",
            "importance": 21
        },
        {
            "uid": "EB092E2F-5B20-DBB0-63D8-26A4898DD944",
            "name": "tacky",
            "importance": 12
```


Funded by the European Commission

```
        },
        {
            "uid": "3CD8FE90-F6BC-B74E-6770-460A4C8DC5F4",
            "name": "rebellious",
            "importance": 24
        },
        {
            "uid": "6305D635-B152-77EA-BD91-6B8336331BA4",
            "name": "active",
            "importance": 30
        },
        {
            "uid": "E50384B0-7F88-3104-39A2-184147A58DE2",
            "name": "communicative",
            "importance": 43
        },
        {
            "uid": "C5131D74-04FA-F960-16D8-49570FB115E5",
            "name": "impulsive",
            "importance": 1
        },
        {
            "uid": "C86494B7-B7E3-1492-419C-9A761C3F03C8",
            "name": "serious",
            "importance": 43
        },
        {
            "uid": "6C998DCC-6FB2-CC67-1848-535B67DE834E",
            "name": "sensual",
            "importance": 43
        },
        {
            "uid": "213BBCF2-F9B1-5AE8-1440-3C8EFAEE7EAF",
            "name": "anxiety",
            "importance": 56
        },
        {
            "uid": "81AE192F-C3B4-2677-7E35-EB342BD95DBF",
            "name": "careless",
            "importance": 24
        },
        {
            "uid": "2E135095-8C4D-34C5-4344-FD60E20718C6",
            "name": "cynical",
            "importance": 21
        },
        {
            "uid": "5E579C46-BD0F-D09B-1B52-498CAE52A095",
            "name": "bravado",
            "importance": 25
        },
        {
            "uid": "06CB5405-5031-9D8B-B099-9B6966AFDB6A",
            "name": "cerebral",
            "importance": 24
        },
        {
            "uid": "2A67558C-2FF3-2D44-6DFB-523F80176453",
            "name": "powerful",
            "importance": 28
        },
        {
            "uid": "63093204-017F-A80B-359E-9B5CC8908878",
            "name": "analytical",
            "importance": 26
```


Funded by the European Commission

```
        },
        {
            "uid": "83E38EEF-C34F-8F75-A789-91BBC6A934E0",
            "name": "revitalizing",
            "importance": 56
        },
        {
            "uid": "B4DF9818-FAB2-528F-A427-4B82652C9BBC",
            "name": " resourceful",
            "importance": 56
        },
        {
            "uid": "5BC6DCE0-5979-2889-25B7-52C05CE2D721",
            "name": "emotional",
            "importance": 50
        },
        {
            "uid": "8E4F875D-7087-64ED-DECA-5295038D9A12",
            "name": "Above (Fire)",
            "importance": 6
        },
        {
            "uid": "6EDA4243-4259-6949-D95E-31F430B9E50F",
            "name": "free",
            "importance": 44
        },
        {
            "uid": "962C3966-5E12-696F-EF7E-68ABF6A2AFA7",
            "name": "Down (Metal)",
            "importance": 17
        },
        {
            "uid": "5FCD7DCD-9C2C-F96D-1EE8-D9A3EF5886AB",
            "name": "sensitive",
            "importance": 99
        },
        {
            "uid": "6CEBAF4C-A524-2EF9-9061-6B5CA6490EA3",
            "name": "pretty",
            "importance": 22
        },
        {
            "uid": "19F1971C-46D1-80DB-659D-7ECA064B86B5",
            "name": "innocent",
            "importance": 33
        },
        {
            "uid": "78AB3FB8-D41B-9369-06A9-2FFC278F9DAB",
            "name": "desirable",
            "importance": 55
        },
        {
            "uid": "CF7D9EA4-6181-7C04-B025-75A9CFED953A",
            "name": "spontaneous",
            "importance": 33
        },
        {
            "uid": "EBD72BDD-08AC-1AF2-B72F-DA30FF41730F",
            "name": "melodramatic",
            "importance": 99
        },
        {
            "uid": "B7004C0A-AB8F-E12B-F1F5-1A8B65867C55",
            "name": "rude",
            "importance": 22
```


Funded by the European Commission

```
                },
                {
                    "uid": "9FCC01A9-1AE4-18BD-B5AE-6CD1C7A1D09E",
                    "name": "romantic",
                    "importance": 77
                },
                {
                    "uid": "044BC3C6-9073-B7F4-E171-31F09E59FD73",
                    "name": "deep",
                    "importance": 30
                },
                {
                    "uid": "993E5DF5-6A3E-EF42-D236-91F34FB8D1CA",
                    "name": "delicate",
                    "importance": 11
                },
                {
                    "uid": "1E66C428-F242-7510-C474-303C768870F4",
                    "name": "dynamism",
                    "importance": 27
                },
                {
                    "uid": "59B0964A-D326-E5B8-22FE-659B160B8EFF",
                    "name": "In/within (Water)",
                    "importance": 17
                },
                {
                    "uid": "F695778E-6475-F312-350C-66114B1BDEC8",
                    "name": "withdrawal",
                    "importance": 1
                },
                {
    ...
```


Funded by the European Commission

5 Conclusions

Deliverable D2.2 “"High-level music content analysis framework v1” is a prototype implementation of the music analysis and indexing framework available to the FuturePulse platform.

This document is an accompanying report describing the prototype implementations.

IRCAM and Musimap provided two prototype implementations.

Future activities for the coming period of the project have the following goals:

- Improve the quality of the automatic estimation of high-level music content analysis. This can be done by improving the machine-learning models used, by increasing the number of training data (either by using data augmentation or by obtaining more data from the Use Case partners) or applying transfer learning from models trained on external data.
- Extend the set of tags considered for each tag family: Electronic-Genre (as defined by BN), tags of Genre (as defined by SYB) and Vocal/Instrumental/Gender tags. This will be done with the help of additional annotations provided by Use Case partners.
- Improve the scalability of the Web Service in terms of increasing the number of simultaneous POST (and then IMD computation) and GET request.

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.