MAC, a system for automatically IPR identification, collection and distribution

[1]Carlos Serrão

[1]ISCTE-IUL/DCTI/ADETTI, Ed. ISCTE, Av. Das Forças Armadas, 1649-026 Lisboa, Portugal
[1]carlos.serrao@iscte.pt

***Abstract*— Controlling Intellectual Property Rights (IPR) in the Digital World is a very hard challenge. The facility to create multiple bit-by-bit identical copies from original IPR works creates the opportunities for digital piracy. One of the most affected industries by this fact is the Music Industry. The Music Industry has supported huge losses during the last few years due to this fact. Moreover, this fact is also affecting the way that music rights collecting and distributing societies are operating to assure a correct music IPR identification, collection and distribution. In this article a system for automating this IPR identification, collection and distribution is presented and described. This system makes usage of advanced automatic audio identification system based on audio fingerprinting technology. This paper will present the details of the system and present a use-case scenario where this system is being used.**

***Index Terms*— digital music, audio fingerprinting, intellectual property rights, piracy, music rights distribution**

I. Introduction

In the digital Era, managing properly and efficiently digital Intellectual Property is an enormous challenge. This is particularly important in a rights distribution scenario that ensures that the appropriate rights are distributed as real as possible to the creators. Such perfect system does not exist at the moment.

Technology plays an important role in the establishment of this distribution scenario and a mechanism has to be established to enable an automated music identification system. This automated system must rely on audio fingerprint technology to automatically identify music tracks from its psycho-acoustic intrinsic characteristics.

This paper presents a method for audio identification on audio streams based on the division of the latter ones in smaller audio elements. Those small parts of audio identified kept in order permit to obtain a more reliable audio identification. The proposed paper will further describe this audio-fingerprinting process, analyze the results obtained in automatic audio identification and introduce an integrated system where it is being tested and used.

The system that was developed to help the correct identification, collection and distribution of the music IPR, covered the different aspects a derived rights collective society has to deal, in order to become more efficient and business centric.

This paper starts by providing an overview of the music rights collecting society processes and describe how these processes are manually handled. From this, a system for automating these processes is presented – the Music Active Control system. A specific technology of the system is presented and described as the technology that enables the system to be fully automated – the audio fingerprinting system. Finally, some details about how the system operates are also provided.

II. Music Related Rights Management Societies

Related Rights (RR) or Neighboring Rights (NR) are terms in copyright law that represent the rights which are similar to the author rights but which are not connected with the actual author of the work. Both the author rights and the related rights are copyrights. The RR/NR is independent of any authors' rights, which may also exist in the work. The rights of performers, phonogram producers and broadcasting organizations are certainly covered, and are internationally protected by the RR/NR legislation. In the specific case of the music industry, and as an example, four different copyright-types rights will concurrently protect a CD recording of a song:

- The authors' rights of the composer of the music;
- The authors' rights of the lyricist;
- The performers' rights of a singer and the musicians;
- The producers' rights of the person or corporation, which made the recording.

Therefore one the most important activities of these Music Related/Neighboring Rights Management Societies (MRNRMS) is the collection of neighboring rights on behalf of producers and performers related to public performance of recorded music.

Therefore the mission of a MRNRMS can be resumed in the following four major objectives:

- Raise public awareness to the reality of related/neighboring rights - which are the rights of artists and producers - and the need for its protection (a fact still relatively new and little known);
- Boosting the delivery of remuneration for distribution to

the holders, be they producers or artists;

- Realize the collection of related/neighboring rights to all places of public performance using recorded music for commercial purposes, as well as all the inspectors to use of recorded music, by any means. The spaces where it is used recorded music for public performance can be clubs, transport systems, hotels, amusement parks and parking, banks, call centers, stadiums, street music, local music and recorded, among others;
- The community awareness in relation to associated rights will, in large part, be accomplished with the collaboration of public authorities with powers of supervision on Copyright and Related/Neighboring Rights, as well as the users of recorded music in various areas and industries that in compliance with the law, should ask for their license.

These MRNRMS are responsible for issuing licenses to businesses that use represented recording music as a mean to conduct their own business models. Moreover, they are also responsible for the effective collection and distribution of the associated fees to the music producers, performers and authors.

Most of this work is accomplished using manual procedures. The verification process is conducted manually as well as the distribution. Moreover the MRNRMS often uses a flat rate fee to charge from their clients. This flat rate fee is calculated based on the type of business and the business space. It is also difficult to know exactly what type of recorded music does the licensee uses – most of the times this evaluation is conducted using a direct sample of music listened during some discrete periods of time. In terms of rights distribution, most of the producers, performers and authors will receive the same amount of money independently if their music was or not used by the client.

This situation creates issues both in terms of rights charging (charging the right price for the music that was effectively used by the client) and in terms of distribution (distributing the right amount of money to the producers, performers and authors whose music was used by the clients). The MRNRMS should pursue the means to be as efficient as possible on the accomplishment of its mission.

In order to solve these problems in terms of collection of fees from the clients, management of the represented repertory and distribution of the producers, performers and author rights, the MRNRMS has to employ the necessary automatic means to become more efficient on its mission.

In the following sections a system that was developed to deal with these issues is presented. The Music Active Control system is a networked based system that empowers the MRNRMS to perform better in the different aspects of their activity.

III. THE MUSIC ACTIVE CONTROL SYSTEM

In order to accomplish its mission in an effective way, the MRNRMS developed a distributed system, based on set of open components and on audio-fingerprinting technology, called Music Active Control (MAC). The MAC system is a web-based system that enables the integration into the MRNRMS value-chain, the different elements that are part of the Related/Neighboring Rights ecosystem. MAC allows the upstream integration of the music producers, and the downstream integration of the music using business clients and the RR/NR proprietary's (Figure 1).

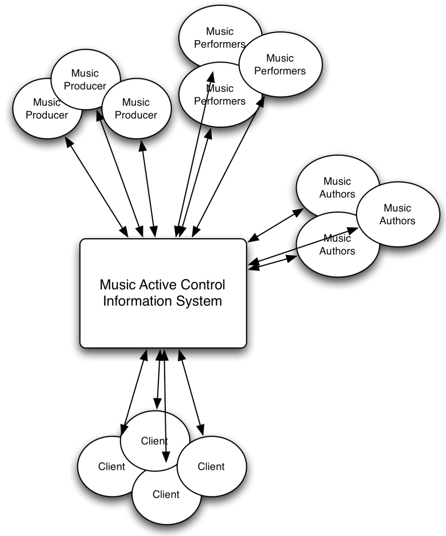

Figure 1. The Music Active Control actor's ecosystem

The MAC system aggregates the information from the music releases from the different music producers, the information of the representative of the music in the specific country (a specific label may be representing this music or artist in the country), information about the performers and information about the music authors.

A. MAC Audio and Metadata Grabber

It is an objective of the MAC system that the different actors themselves should update the data on the system – this way it will be possible for the system to grow with little central coordination and control effort. In order for this objective to be achieved, the MAC system uses a web-based portal as well as a client-side application (Figure 2) to feed the MAC system database with the necessary data.

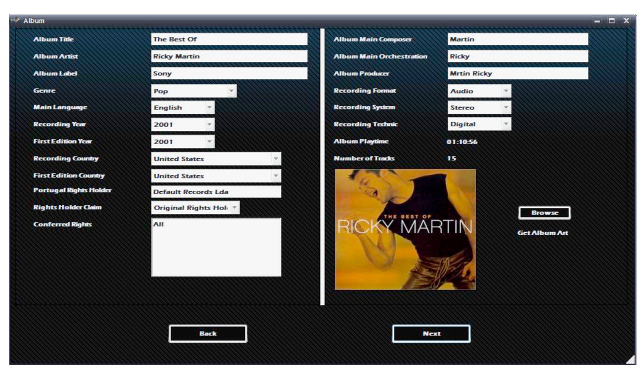

Figure 2. Client application that allows the data entry on the MAC system

The client side application, the MAC Audio and Matadata Grabber that is distributed to the music producers, allows an easy and seamless way to capture the MAC system required data. This application, not only captures the information from the new music releases, but also crucial information about the audio music itself. The audio data is analyzed and extracted from the audio CD, high-quality MP3 files are created and an audio-fingerprint is generated from the original audio CD content. The full process can be described on the following steps:

- The producer starts the application;
- The producer inserts an newly released audio CD on an optical computer drive;
- The application analyses the audio CD data, and tries to extract metadata both from the audio CD and from several Internet databases. This makes the data filling task less complex and faster;
- The producer verifies the metadata that was automatically from the audio CD or the Internet, corrects it and or completes it – there are some information that is impossible to extract directly from the Internet and must be manually input;
- The application analyses the audio CD data, and generates a unique fingerprint for each of the music tracks that are part of the CD;
- The application extracts the audio from the CD and converts it to an high-quality MP3 file;
- Finally, the application sends all the metadata, MP3 files and the audio fingerprints to the MAC system.

The MAC system stores all the information on a centralized database and indexes the MP3 music files. Also, the audio fingerprints are associated to the music track metadata, for later matching processes. When this process completes, the MAC system contains the information about all the different music tracks that the MRNRMS represents, and that require licensing from businesses using that music for commercial purposes.

B. MAC.box

The other important process of the MAC system is the automatically identification system that will be used at the client business side, to collect the exact information of the music that was used by the client, in order for the MRNRMS to charge the appropriated fee and to correctly distribute the RR/NR rights associate to the music that was used.

In order to accomplish this, the Mac system makes available on the client side, a small hardware/software appliance, called MAC.box. This appliance is responsible for recording the audio data, from the used audio source, and to perform the generation of audio fingerprints. From time to time, the MAC.box connects to the MAC system and sends the captured information. This information is matched on the MAC system side, and the represented music that has been used by the client is automatically identified.

The identification and matching process can be described in the following steps:

- The MAC.box is connected directly to the output of the audio device that will be playing music on the client business;
- The MAC.box captures (“listens”) the audio being played by the audio device, on specific time intervals and for short periods of time, recording the captured data;
- MAC.box computes the audio fingerprints of the different samples captured;
- The MAC.box sends a report containing the list of computed fingerprints, and the date and hour of capture to the central MAC system;
- The MAC system, compares the received captured fingerprints against the original fingerprints that are stored on the system;
- If both fingerprints are a match the music is identified;
- The final result of this process is a series of reports that list the music tracks that were used by a certain client, during a certain period of time.

With this information, the MRNRMS can charge the customer better and perform a fairer distribution of rights to producers, performers and authors.

All of the processes described related to the identification of the music tracks are only possible because of a key-enabling technology: the audio-fingerprinting system. In the following sections, a more detailed description of the audio-fingerprinting system used on the MAC system for music matching, is described.

IV. The MAC system and Audio-Fingerprinting

As it was previously referred on this paper, the system is based on the automatic identification of the different music tracks that are used by the clients. For this automatic music identification to work, two different processes are necessary in the system (Figure 3):

1. The creation of the original audio-fingerprint from the original CD music tracks and their storage associated with music metadata;
2. The comparison between an audio-fingerprint candidate (or multiple audio-fingerprint candidates) and the original audio-fingerprint stored on the system. If a match is found than the music is identified and the corresponding metadata is pulled from the system.

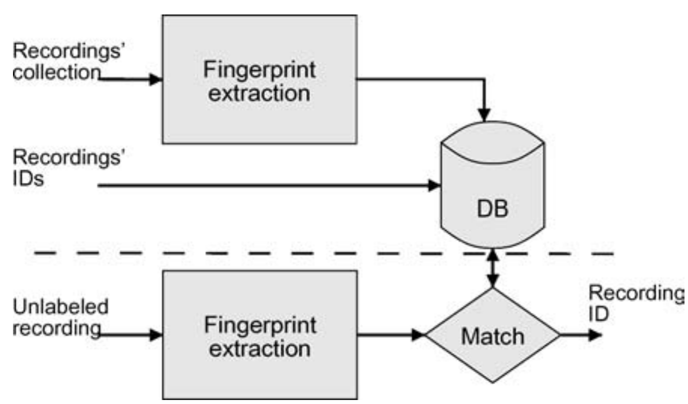

Figure 3. Processes involved in the generation and matching of audio-fingerprints

Both these processes rely on the capability of the audio-fingerprinting algorithm for correct audio identification. The audio-fingerprint or acoustic fingerprint is a condensed digital summary, deterministically generated from an audio signal that can be used to identify an audio sample or quickly locate similar items in an audio database.

V. AUDIO FINGERPRINTING BASED ON SVD

Any complete audio identification system based on audio fingerprinting technology is normally divided in two distinct parts. First, an audio fingerprint extraction process that is responsible for extracting an unique identifier (fingerprint) from the music track and second, the matching process that tries to find a match between the generated fingerprint and the different fingerprint identifiers stored on a database and the corresponding meta-information.

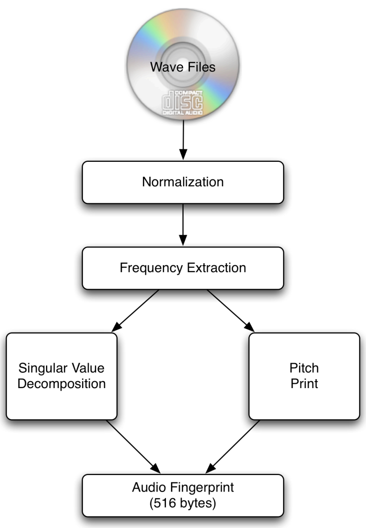

Figure 4. Description of the process for the creation of audio-fingerprints

Considering the specific requirements of the system to be developed, there is a set of requirements that the audio-fingerprint generation process [1] [2] should be able to comply to:

- **Robustness**: precise identification of the audio, despite the existence of signal degradation. The same music with different qualities should generate the same audio fingerprint. To identify the robustness of a system, it is normal to use the rate of false negatives, which is the rate of audio fingerprints that the system can not detect, despite the information for their identification is in the repository;
- **Efficiency**: the ability to correctly identify the maximum possible of music and recognize the associated information. Usually to describe the efficiency of a system the rate of false positives is used, which is the rate of incorrect identification of audio fingerprints;
- **Granularity**: the ability to identify audio samples with just a few seconds duration. This implies working with synchronization between audio fingerprints taken and those that are stored in the repository. As this parameter depends on the objectives of the system in some situations it may be necessary to use the entire song instead of just an excerpt;
- **Dimension**: for quick searches, it is necessary to store the audio fingerprint in RAM. Consequently, the memory resources to the server database will have to be determined taking into account this parameter;
- **Scalability**: the ability to find an audio fingerprint in a reasonable time, taking into account the number of existing audio fingerprint in the repository. The larger the database, the greater the physical capacity of storage required. It should be possible to resize the system with ease, due to the high growth potential of these systems.

The production of an audio-fingerprint with the about requirements is not an easy task. It involves a different set of tasks from the audio source, until the final fingerprint is computed [3].

The pre-processing of the audio signal is performed in order to prepare it for the following analysis, carried out by other components.

As a first step the audio is passed to mono format (in the case of a stereo signal), the silence is removed, followed by removal of DC offset, given that its existence does not allow to maximize the volume during the normalization process.

The pre-processing finishes by converting the signal to 44.100Hz. Finally, the 10 initial seconds of music are removed in order to avoid the silence that could have been introduced in the audio. The remaining audio signal after the initial 120 seconds is ignored, allowing larger music coverage and preventing a file manipulation attack [4].

After the normalization of the signal the frequency is extracted using the Fast Fourier Transform (FFT).

Upon receiving the data returned by the FFT, the spectrum series are examined, each one representing a small audio frame (185 milliseconds). In turn, each frame is composed by a set of frequencies (frequencies matrix).

The system proposed on this paper will use the Singular Value Decomposition (SVD) applied to the matrix of audio frequencies for the extraction of audio fingerprints, which produces a smaller size matrix result resuming the audio features [5]. In order to represent an audio element $\boldsymbol{X}$ as a matrix, it will have to be divided in $\boldsymbol{M}$ audio segments, each of them containing $\boldsymbol{N}$ frequency bins, originating therefore a matrix $\boldsymbol{M*N}$, where $\boldsymbol{M>N}$. With the aim of obtaining a more reduced matrix is applied to the matrix $\boldsymbol{X}$ the SVD, decomposing it as

$$X = USV^T$$

where $\boldsymbol{U}$ with size $\boldsymbol{M*M}$ and $\boldsymbol{V}$ with dimension $\boldsymbol{N*N}$ are orthogonal matrices and their columns are called respectively left and right singular vectors. $\boldsymbol{S}$ is a diagonal matrix of singular values. These singular values are sorted in a decreasing order and represent the importance of the correspondent singular vectors in the matrix structure. In general the first singular values contain most part of the information relative to the matrix, thus in order to reduce redundant information it is advised to only keep primary singular values. In the proposed system, it was decided to keep four singular values, after the analysis of experimental results. Those singular values together with the correspondent right singular vectors, which contain spectral principle components, can be used to represent the main features of the audio element. In the system presented on this paper they are used as part of the audio fingerprint that has been called CorePrint. The other part is composed by four numbers that correspond to the most prominent frequencies of the audio pitch and is called PitchPrint. PitchPrint is useful for the lookup and matching procedures [6] [7].

At the end of the fingerprint extraction process, a 408 bytes fingerprint (8 bytes from the Pitch, 320 bytes from the right singular vectors and 80 bytes correspondent to the singular values) resuming the audio element analysed, is obtained.

The matching process is based on a first part possible results reduction, using a pitch match procedure. The obtained results are produced by a more complex and time consuming matching process in order to obtain the identification to each audio fingerprint. After this, the outcomes are analyzed as a whole, and the decision about which music is being played is made. The foremost part of the matching procedure (pitch match) is a very simple one, consisting in the direct comparison of the audio element pitch numbers analyzed with the ones stored on a database. This way all the fingerprints that do not have equal PitchPrint will be discarded as possible match. With the possibilities reduced, the system will aim at the CorePrint match. For that it is necessary to compute the similarity value between each print, as defined here

$$S = \sum_{i=1}^{m} \sum_{j=1}^{m} w_{i,j} s_{i,j}$$

where ***Wi,j***

$$w_{i,j} = \frac{\lambda_{c1,i} \lambda_{c2,j}}{\sqrt{\sum_{i=1}^{m} \lambda_{c1,i}{}^{2} \sum_{j=1}^{m} \lambda_{c2,j}{}^{2}}}$$

is the variable weight.

This specifies how each ***Si,j*** will contribute to the similarity calculation, and ***Si,j*** is the dot product between two groups of ***S*** vectors (right singular vectors). The higher similarity will correspond to a match and the corresponding audio identification. This is only valid if its value is superior to a pre-defined threshold, therefore reducing the false positives probability. When an audio stream is divided in audio elements and those are analysed, a set of audio elements identifications will be produced. Setting the size of the audio elements to a relative small size comparing with the normal size of a music, and maintaining the output results ordered will probably originate consecutive equal identifications [8]. The system presented on this paper will consider that music has been correctly identified when a defined number of equal consecutive audio elements identifications are found.

VI. Tests and Results

In order to assess the quality of the system and in particular the robustness of the audio-fingerprint some tests were conducted. In this assessment a set of different music tracks (from different artists and different albums') were selected (20 audio music tracks), and three different tests were conducted to evaluate this robustness.

The first test aimed to confirm the ability of the audio-fingerprint to work with different sampling patterns. Each file was converted to a lossy MP3 format, and for different sampling frequencies (32, 44 and 48 kHz) and sampling speeds (32, 48, 64, 96, 128, 160, 192 kbps). The following table (Table 1) summarizes the results for the different sampling rates, taking into account that this factor did not alter the results.

Table 1. Testing the Different Rhythms and Sampling Frequencies (P: Passed; F: Fail; NT: Not Tested)

Artist	Music	32kbs	48kbs	64kbs	96kbs	128kbs	160kbs	192kbs
Green Day	Nice Guys Finish Last	P	P	P	P	P	P	P
Green Day	Hitchin'A Ride	P	P	P	P	P	P	P
Green Day	The Grouch	P	P	P	P	P	P	P
Green Day	Redundant	P	P	P	P	P	P	P
Green Day	Scattered	P	P	P	P	P	P	P
Smashing Pumpkins	I am one	P	P	P	P	P	P	P
Smashing Pumpkins	Siva	P	P	P	P	P	P	P
Smashing Pumpkins	Rhinoceros	P	P	P	P	P	P	P
Smashing Pumpkins	Bury me	P	P	P	P	P	P	P
Smashing Pumpkins	Crush	P	P	P	P	P	P	P
Queen	Keep yourself Alive	P	P	P	P	P	P	P
Queen	Doing All Right	P	P	P	P	P	P	P
Queen	Great King Rat	P	P	P	P	P	P	P
Queen	My Fairy King	P	P	P	P	P	P	P
Queen	Liar	P	P	P	P	P	P	P
Silence 4	Goodbye Tomorrow	P	P	P	P	P	P	P
Silence 4	Borrow	P	P	P	P	P	P	P
Silence 4	Dying Young	P	P	P	P	P	P	P
Silence 4	Old Letters	P	P	P	P	P	P	P
Silence 4	Angel Song	P	P	P	P	P	P	P

As shown by the results obtained, the audio fingerprint system worked with all the rhythms and frequencies of sampling with which it was tested.The aim of this second test was to test the ability of the audio fingerprinting to deal with the introduction of silence at the beginning of the audio files. The following table (Table 2) summarizes the results obtained.For this second test has been included at the beginning of each of the audio files, 5, 10, 15 and 20 seconds of silence. The audio music files used on the test were in MP3 format with sampling frequency of 44kHz and 128kbs sampling rate (typical values for audio files). After this, the fingerprints were generated for the music audio files containing the silence and analyzed the results by comparing with the fingerprints of the original unmodified audio files. As shown by the results table (Table 2) obtained from the test the ability to fight the introduction of silence at the beginning of the audio is yet not ideal.

Table 2. Testing the introduction of silence at the beginning of the audio music files (P: Passed; F: Fail; NT: Not Tested)

Artist	Music	5s	10s	15s	20s
Green Day	Nice Guys Finish Last	P	F	F	F
Green Day	Hitchin'A Ride	P	P	F	F
Green Day	The Grouch	P	P	F	F
Green Day	Redundant	F	P	P	F
Green Day	Scattered	P	F	F	F
Smashing Pumpkins	I am one	F	P	P	F
Smashing Pumpkins	Siva	F	F	F	F
Smashing Pumpkins	Rhinoceros	NT	NT	NT	NT
Smashing Pumpkins	Bury me	P	F	F	F
Smashing Pumpkins	Crush	P	P	F	F
Queen	Keep yourself Alive	P	F	F	F
Queen	Doing All Right	F	F	F	F
Queen	Great King Rat	F	F	F	F
Queen	My Fairy King	F	F	F	F
Queen	Liar	F	F	F	F
Silence 4	Goodbye Tomorrow	NT	NT	NT	NT
Silence 4	Borrow	NT	NT	NT	NT
Silence 4	Dying Young	P	F	F	F
Silence 4	Old Letters	F	F	F	F
Silence 4	Angel Song	F	F	F	F
	Passaram	8	5	2	0
	Falharam	9	12	15	17

In this third test, as before, were used MP3 files with a sampling frequency of 44kHz and 128kbs rhythm sampling.

Table 3. Simulation of AM reception and introduction of distortion (P: Passed; F: Fail; NT: Not Tested)

Artist	Music	AM Effect	2x AM Effect	0,5 Distortion	1,0 Distortion
Green Day	Nice Guys Finish Last	F	F	P	P
Green Day	Hitchin'A Ride	P	P	P	F
Green Day	The Grouch	P	F	P	F
Green Day	Redundant	F	F	P	F
Green Day	Scattered	P	F	P	F
Smashing Pumpkins	I am one	F	F	P	F
Smashing Pumpkins	Siva	P	F	P	F
Smashing Pumpkins	Rhinoceros	NT	NT	NT	NT
Smashing Pumpkins	Bury me	P	F	P	F
Smashing Pumpkins	Crush	P	F	P	F
Queen	Keep yourself Alive	P	F	P	F
Queen	Doing All Right	P	F	P	P
Queen	Great King Rat	F	F	P	F
Queen	My Fairy King	F	F	P	F
Queen	Liar	F	F	P	F
Silence 4	Goodbye Tomorrow	NT	NT	NT	NT
Silence 4	Borrow	NT	NT	NT	NT
Silence 4	Dying Young	F	P	P	F
Silence 4	Old Letters	F	F	P	F
Silence 4	Angel Song	F	F	F	F
	Passaram	8	2	16	2
	Falharam	9	15	1	15

In this final test errors were introduced in the audio signal such as distortion or effect caused by bad signal reception (different levels of signal distortion). The results (Table 3) shown that the audio fingerprinting can correctly identify the music if the distortion rate is low. When the distortion rate increases the identification capability diminishes considerably. However this was not one of the biggest requirements of the system, because we will have direct access to the audio device output with low distortion.

VII. Conclusions

This paper presented a system that empowers the Music Related/Neighboring Rights Management Societies (MRNRMS) to perform much better. This reflects in the way the clients are charged and on the way the different related or neighboring rights are distributed to the producers, performers and authors. The systems uses the new information and communication technologies, in a distributed system, over the Internet, to create and collect information about the managed music portfolio, and for the collection of information about the music consumption on the business clients. The music portfolio is updated by the music producers through the automatic introduction of metadata, and by the generation of unique audio identifiers – the audio/acoustic fingerprint – that will enable the automatic identification of the different audio tracks. At the business client side, the MRNRMS relies on a set of small appliances (MAC box) that is connected to the audio output of the audio device, and that captures small samples of the audio in discrete periods of time, generating a set of candidate fingerprints. These candidate fingerprints, as well as some additional capture data is sent over the Internet, to the MAC system, for the matching process. On the MAC system, the different audio-fingerprint candidates will be searched on the database against a similarity threshold that will allow the identification of the audio music track that was played. As such, the music identification processes that were previously perform using a manual procedure can now be completely automated and the system brings important performance gains for the operation of MRNRMS.

References

[1] Burges, C. J., Platt, J. C., & Goldstein, J. (8 de Novembro de 2003). Identifying Audio Clips with RARE. Berkeley, California, U.S.A.

[2] Burges, C. J., Platt, J. C., & Jana, S. (2002). EXTRACTING NOISE-ROBUST FEATURES FROM AUDIO DATA. International Conference on Acoustics, Speech, and Signal Processing, (pp. 1021-1024). Redmond.

[3] Cano, P., Batlle, E., KALKER, T., & HAITSMA, J. (2005). A Review of Audio Fingerprinting. Journal of VLSI Signal Processing , 41, 271–284.

[4] Doets, P., & Lagendijk, R. Stochastic Model Of A Robust Audio Fingerprinting System. Delft.

[5] Haitsma, J., & Kalker, T. (2002). A Highly Robust Audio Fingerprinting System. IRCAM.

[6] KLEMA, V. C., & LAUB, A. J. (2 de Abril de 1980). The Singular Value Decomposition: Its Computation and Some Applications. IEEE TRANSACTIONS ON AUTOMATIC CONTROL, , pp. 164-176.

[7] Lu, L., & Hanjalic, A. (27 de Outubro de 2006). Towards Optimal Audio "Keywords" Detection for Audio Content Analysis and Discovery. Santa Barbara, California, USA.

[8] MusicIP. (2006, Março 12). Open Fingerprin Architecture WhitePaper v1. Monrovia, California, U.S.A.