DESCRIBING ACOUSTIC FINGERPRINT TECHNOLOGY INTEGRATION FOR AUDIO MONITORING SYSTEMS

Carlos Serrão
carlos.serrao@iscte.pt
ADETTI/ISCTE
ISCTE - Instituto Superior das Ciências do Trabalho e da Empresa
1600-082 Lisboa, Portugal

Marco Clara
antunes.clara@marinha.pt
Direcção de Tecnologias de Informação e Comunicação
Marinha Portuguesa
1149-001 Lisboa, Portugal

***Abstract* - The ability to create, write or compose a song is a gift or feature that not everyone has within him or herself. The results of that kind of ability are an asset that should somehow, due to their nature, be recorded and protected. That protection can be enabled by the enforcement of the author's rights, safeguarding its works. How can this protection be enforced in a digital World? Should we trust that nobody steals or circumvents intellectual or artistic property? Or should we think of possible ways to control and monitor that property's usage? It seems that the last one is currently the right way, and technologies such as acoustic fingerprint may allow us to provide such monitoring and enforcement. In what way can we integrate existing technology for creating such systems, and what criteria should be used for evaluating that same technology?**

***Index Terms* - DRM, Digital, Rights, Restrictions, Management, Audio, Acoustic, Fingerprinting**

I. INTRODUCTION

Music is practically a constant in our day-to-day life. Imagine this possible following daily scenario: we wake up in the morning, and listen to the radio station that is playing our wake up song. Some of us even use a small radio in the bathroom so that we can hear it while taking a bath. We drive our cars to the job, still listening to our favourite radio, our favourite CD or a MP3 compilation of our favourite songs. We arrive at work, and plug in our headphones so we can keep listening to music. We go to lunch and the restaurant is playing ambience music. We get back to work and to our headphones. We go back home, driving our car and still listening to music. We arrive home and turn on our audio system or start watching a concert in our DVD player. If we go out to a bar, or a disco, music is still there. In practice we can stay all day long consuming music from several different sources and not even thinking about it – it is something that has become "natural" in our lives. We may not even be asking for it, and yet, it's there. So what about the music creators/authors? Who's taking care of their interests or their rights? If we buy a CD or a DVD (disregarding copy protection issues) their rights are safeguarded, but what about all the other situations? The resulting need for systems that can enforce some type of protection and monitoring is becoming more and more important and several technologies are now emerging so that these systems can be implemented. We will start by giving some ideas about some of the most common systems used.

Digital Rights Management (DRM) [1] is the concept that refers to a set of several different technologies that allow enforcing control policies, in what concerns digital data (software, hardware), video and audio. The monitoring of the rights, also enclosed within DRM concept, is probably one of the most difficult tasks to enforce, and it deserves some attention.

However, one should not be confused about this concept (DRM) and others such as "copy protection". In this case we are talking about technologies that simply allow controlling the access, or restricting the usage of digital media, according to the used device for playing it, for example. Some critics claim that a more accurate designation for DRM should be Digital "Restrictions" Management [16]. But, if we start using the "restrictions" term, we are somehow limiting DRM concept, in the sense that, for example, the usage monitoring of contents for rights recording is also part of DRM enforcing. Anyway, the main use of DRM related technologies is considered to be the revenue loss prevention due to illegal duplication of copyrighted contents [17], although this is not the main issue for this document.

According to the previously mention needs concerning enforcement of DRM policies, and specifically, to the monitoring ability of content usage, it is important to clarify some aspects related to a specific technology. This paper is focused primarily on audio content monitoring and protection, the acoustic fingerprint (also called audio fingerprint).

The main objective of an acoustic fingerprint mechanism is to efficiently and robustly compare the equality (or not) of two audio files, not by comparing the files themselves, but by comparing substantially smaller sets of information, referred to as acoustic fingerprints [8]. Usually, systems relying on acoustic fingerprinting, all the audio file information, including the file itself, artist, title, album duration or any other kind of related information is maintained in some kind of database, being the fingerprint the main identifier (or index) for the record. This way, when searching a record for comparison, one may obtain more quickly and robustly a result, either if there is a match or not in the database. In result, the fundamental aspects of this technology can be identified:

- Fingerprint generation mechanism
- The fingerprint itself (previously generated)
- The fingerprint matching mechanism
- Database for audio related information (including the fingerprint as index)

This paper will describe in further detail the acoustic fingerprinting technology, not only by covering the previously mentioned fundamental aspects or components, but also considering integration aspects related to usage of this technology on other information systems. This way, we intend to provide some information that can be used as a set of guidelines for analysing related technology and

T. Sobh (ed.), Innovations and Advanced Techniques in Computer and Information Sciences and Engineering, 139–144.

implementation mechanisms, regarding audio monitoring systems.

II. TECHNOLOGY

A. Description

An acoustic fingerprint can be considered as some kind of DNA (deoxyribonucleic acid) scheme for the associated audio file or object. This way we can say that, like DNA, the acoustic fingerprint of any audio file will be unique. We can also associate the concept of hash functions to acoustic fingerprinting, since a hash function F should map a large object X to a smaller corresponding hash value. So, we can think about comparing hash values for two submitted objects for the same hash function. This way, if the obtained hash value is correspondent, then the submitted objects should in fact be the same. The only difference between a simple hash function and an acoustic fingerprinting algorithm is that when we think of the objects they may not in fact be the same. Today's audio formats allow us to possess the same audio track with different features, in what concerns, for example, bitrate or other aspects regarding audio quality. So, this means that an acoustic fingerprinting algorithm must be based on the perceptual features of an audio object, and not the "physical" features of the object itself (for example, binary code), feature that obviously falls out of a simple hash function scope.

Despite the mentioned perceptual features of the audio files to be analysed, one should distinguish between a song with varying quality, and two versions of the same song (for example, sung by artist A and sung by artist B). In this case it can be said that such kind of correspondence is virtually impossible to attain. Even though some work is being done in what concerns constructing a "fingerprint function in such a way that perceptual similar audio objects result in similar fingerprints" [8], as described by Jaap Haitsma and Ton Kalker, from Philips Research.

One of the aspects related to an acoustic fingerprinting algorithm is its robustness. We will see in the Open Fingerprint Architecture section (see Section III.B) that to obtain the right robustness, the algorithm should be based on certain perceptual features of the audio, so that even with degradation (radio transmission, for example) the audio signal should still be correctly identified. This is possible through the analysis of invariant features of that same audio that will lead (hopefully) to generating the same fingerprint for different quality (but same audio) content. Haitsma and Kalker [8] also refer to aspects such as "reliability" and "granularity" as fundamental, but as you will see these aspects are indeed mainly related to the robustness of the acoustic fingerprinting algorithm itself. As for the "fingerprint size", "speed" and "scalability" issue, it is all a matter of resources and coding. Since the fingerprint data itself is not that large, storage is practically not a problem. This way, robustness is in fact the main issue when it comes to acoustic fingerprinting, even though all the other mentioned aspects should still be considered very important.

B. Client Applications

One of the most desired uses for acoustic fingerprinting, like mentioned before, is the creation of equally robust monitoring mechanisms that can somehow allow the enforcement of some DRM principles. A simple target of such mechanisms is most definitely the **radio stations broadcasting**, that could this way be precisely monitored, registering every single song played during a certain period. That would allow a more effective and efficient digital rights management and application. Actually, in several countries the best that rights management societies can do is some kind of statistic treatment of the data related to these broadcasts, using samples to calculate which songs are played, with low degree of accuracy. This makes obvious all the advantages that the monitoring systems would bring to this area, giving a more precise feedback regarding multiple channel audio broadcasting. Through this mechanism there would be automatically generated playlist reports for a simpler royalty collection, instead of the mentioned manual process of determining which songs are more or less played in that radio or web station, for example.

Further in this paper it is possible to understand the server technology regarding audio fingerprinting services and data storage, but what about **monitoring systems**? They are one of the perfect candidates to a correctly elaborated plan in what concerns the design of client applications. One can think of a system such as this, as a separate part of the server's architecture. They can take form, for example, of some kind of intermediate appliance between the fingerprint server and the transmission in course. A monitoring appliance could this way be "listening" to a broadcast, registering all the played songs, and automatically identifying and recording those same songs in a log for further future analysis. This way we open the door to a simple and integrated implementation recurring to existent systems and data (or metadata) repositories, making use of (in a perfect world) large sets of information for music identification.

However, we do not live in such a perfect world, and the radio station broadcasting is also the perfect example for the difficulties associated with monitoring systems: the **audio quality**. This issue is a sensitive one, since it makes necessary to have a very robust audio fingerprinting mechanism that can support all the quality degradation present on an analogical signal radio transmission (just think of the difference between the quality associated with AM and FM radio stations and frequencies).

C. Server Applications

It is fair enough to say that, beside algorithms, techniques related with fingerprint generation or client applications of any kind, the remaining complexity is fairly attributed to server side technology. It is possible to understand that, due to the complexity and purposes of a server that delivers data or metadata related to music or song records, more precisely to scalability details, performance requirements, storage capability, etc., this is a crucial point of any adopting architecture. MusicIP [5] is one of the currently greatest metadata and acoustic fingerprint providers in the market, and is also the owner of the repository containing one of the largest sets of information available. This organization is described in further detail on the next section. According to a recent announcement from MusicIP, "the company's global music