Computer and Information Science; Vol. 6, No. 1; 2013
ISSN 1913-8989 E-ISSN 1913-8997
Published by Canadian Center of Science and Education

Efficient Audio Fingerprint Application Verification Using the Adapted Computational Geometry Algorithm

Marios Poulos[1], Ioannis Deliyannis[2] & Andreas Floros[2]

[1] Dept. of Archiving and Library, Ionian University, Corfu, Greece

[2] Dept. of Audio and Visual Arts, Ionian University, Corfu, Greece

Correspondence: Marios Poulos, Dept. of Archiving and Library, Ionian University, Corfu GR-49100, Greece. Tel: 30-266-108-7429. E-mail: mpoulos@ionio.gr

Received: December 2, 2012 Accepted: December 27, 2012 Online Published: January 10, 2013

doi:10.5539/cis.v6n1p70 URL: http://dx.doi.org/10.5539/cis.v6n1p70

Abstract

An earlier work of the authors introduced an adapted version of the Computational Geometry Algorithm (CGA) designed to analyse an audio stream and produce a unique coding-independent fingerprint. As the adaptability and the induced calculation load of the proposed algorithm form a key characteristic for multiple applications, our current investigation aims to measure its performance and stability in dynamic, real-time applications, i.e., in large audio library indexing and dynamic audio recognition. In addition, we investigate the fact that context similarity is also evident across fingerprints; hence a number of comparisons are used to explore the possible uses of this highly desirable algorithmic feature.

Keywords: audio fingerprinting, computational geometry, audio signal processing, experimentation

1. Introduction

The wide availability of Internet connectivity and media-enabled devices has altered the way content is produced, distributed and finally reproduced. Increased demand for all types of audiovisual streams is clearly evident, as most network traffic today consists of multimedia data exchanged in global scale (Deliyannis, 2012). Digitization and networked distribution of audio, video, live broadcasts and the increased demand for interactive control, leads authors and companies to the path of content re-use and reproduction via customization of existing content and distribution (Karydis, Deliyannis, & Floros, 2011) through networked multimedia databases and multicast systems. Within evolving markets such as electronic multimedia-content retail and exchange points, various new services emerge (Deliyannis, Karydis, & Anagnostou, 2011). In comparison to traditional media, new access methods alter the way that data are distributed and reproduced, often forming new applications and domains such as interactive and new-media arts (Trifonova, Jaccheri, & Bergaust, 2008), while changing the user culture in terms of content use (Gillespie, 2004). These global changes introduce new markets and services (Simpson, 2004), a fact that is clearly evident when observing the evolution of standards such as MPEG-7 linking content to context and offering multimedia accessibility for all as with MPEG-21 (Kosch, 2004). The considerable content availability through various media will certainly require new broadcast-control and verification mechanisms to be established, a sector that may be aided by current research.

Computers are also employed in the area of copyright management under a wide variety of applications, one example being the application of pattern-matching algorithms and techniques to identify copyrighted content (Furht & Kirovski, 2005; Karydi, Karydis, & Deliyannis, 2012). The task is straightforward in text-based applications. This particular data format is transferred and delivered in complete form, without loss of content during transfer and reproduction, a fact that significantly aids the pattern-matching process. In contrast, audio (and/or video streams) are often degraded in terms of quality due to the employment of various compression techniques and the inevitable stream re-compression processes introduced by the wide variety of transmission formats available in all media-enabled platforms. These algorithmic-based compression processes such as MPEG 1, 2, 3 and 4 are based on mechanisms of human perception for minimizing the required transmission bandwidth. Ultimately, conversion between various media formats alters significantly the original information, a fact that introduces various problems in the identification process, as error and distortion are clearly evident when contrasting original versus transmitted data.

The present paper can be considered to be a continuation of our latest research which lays the necessary theoretical foundation research on audio fingerprinting based on convex layer definition in the frequency domain (Poulos, Deliyannis, & Floros, 2012). In this work the experimental aspects of a novel algorithm for defining convex layer areas over audio signal spectral peaks as a track identification procedure are addressed in an attempt to standardize the identification process. According to our view, the latter process is clearly identified as a key issue that needs to be resolved before this technology may be exploited commercially. Our experimentation indicates that beyond direct pattern matching, dynamic content detection is also possible. In that respect, once the standards are related to the semantic layers (fingerprints) of information and communication systems, important consequences arise that require further research under Music Information Retrieval (MIR) research (Aucouturier & Pachet, 2003; Casey et al., 2008; Chandrasekhar, Sharifi, & Ross, 2011; Levy & Sandler, 2009; Logan, Ellis, & Berenzweig, 2003; Marsden, 2010; McFee, Barrington, & Lanckriet, 2010; McKay & Fujinaga, 2008; Slaney, Weinberger, & White, 2008; Wang, 2003). Our previous research indicates application areas such as gaming (Deliyannis, Karydis, & Anagnostou, 2011; Karydis et al., 2011) and copyright identification (Deliyannis, Karydis, & Karydi, 2011; Karydi et al., 2012).

The paper at hand is organized as follows. Section 2 briefly presents a synopsis of the Computational Geometry Algorithm (CGA) audio fingerprinting algorithm. This is an issue covered fully in our theoretical definition of the above algorithm published recently (Poulos et al., 2012) and the reader is encouraged to refer to this article for a detailed algorithmic and mathematical analysis. Next, Sections 3 and 4 provide extended experimentation cases based on a number of widely-employed application scenarios and present the results obtained using multiple forms of audio content as well as the statistical evaluation of the derived data. Finally, Section 5 concludes this work by proposing future research directions.

2. Related Study

In our latest study, a novel audio content identification (matching) approach is presented, based on the significant reduction of the original spectral peaks enclosed in convex layer areas (Poulos et al., 2012). This work introduced audio-track identification through the use of computational geometry algorithms, where the problem of matching sample peaks with original peaks was addressed using an intersection technique between convex layers. In particular, this approach produced a convex polygon in the frequency domain that resembles a coordinate-based pattern in terms of a unique set of points that can be considered to be the audio data "fingerprint." In the above work it was also shown that this fingerprint pattern is coding-independent, a fact that provides indications that the proposed algorithm may be suitable for multiple purposes and applications, including the categorisation of content identity and the identification of audio clips, hence providing support for the realisation of audio sorting/searching tasks and services.

The above described method was realised via the use of the Computational Geometry Algorithm (CGA), a computationally efficient scheme of onion-like layers that results into unique frequency-domain representations of the innermost onion layer (Poulos et al., 2012). More specifically, the digital audio signal under identification (test signal), denoted here as $x(n)$, is initially transformed in the frequency domain and represented in terms of its Power Spectral Density (PSD) $X(f)$ via Bartlett's estimation. The same procedure is applied on the original (reference) signal $x_{ref}(n)$, producing the $X_{ref}(f)$ PSD vector of size N. Then, the CGA algorithm is applied on the derived PSD data, producing onion-like layers denoted in the case of reference signal as S. An example of such algorithmically constructed layers is graphically represented in Figure 1. Finally, a critical algorithmic parameter, the total depth of layers (or the k-depth value) is defined, following the algorithm described again in our latest study (Poulos et al., 2012). Finally, by algorithmically isolating the k-th inmost layer, we obtain the convex subset S_{xy} that corresponds to the reference signal. The same procedure is applied on the test signal PSD data and the k-th convex subset N_{xy} is similarly derived. During the final matching/identification process, the intersection of the above convex subsets S_{xy} and N_{xy} is computed, that is:

$$S_{xy} \cap N_{xy} = R_{xy} \tag{1}$$

The identification procedure is completed by extracting the degrees of similarity s_1 and s_2 using the computed areas (A) of the calculated convex subsets (S_{xy}, N_{xy} and R_{xy}) using the following fractions (see also Figure 1):

$$s_1 = \frac{AR_{xy}}{AS_{xy}} \qquad s_2 = \frac{AR_{xy}}{AN_{xy}} \tag{2}$$

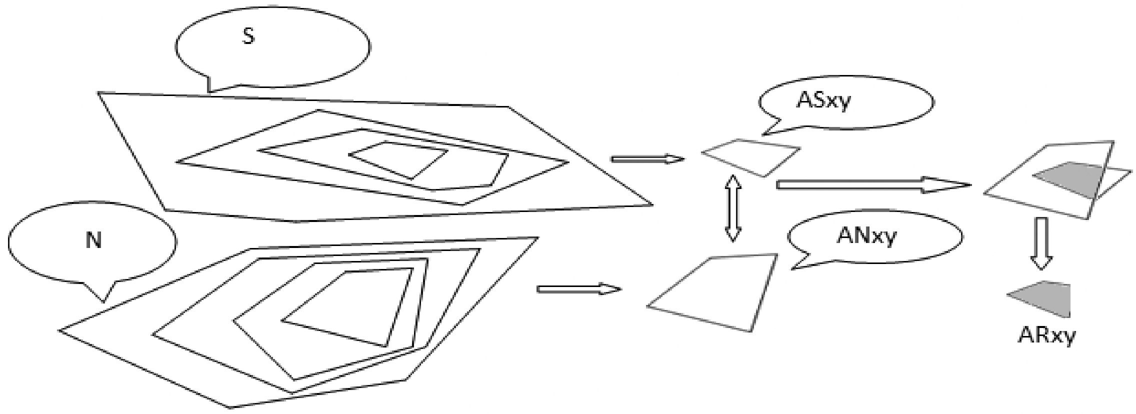


Figure 1. A graphical representation of the onion-like layer extraction process

The above identification/matching process architecture is graphically illustrated in Figure 2.

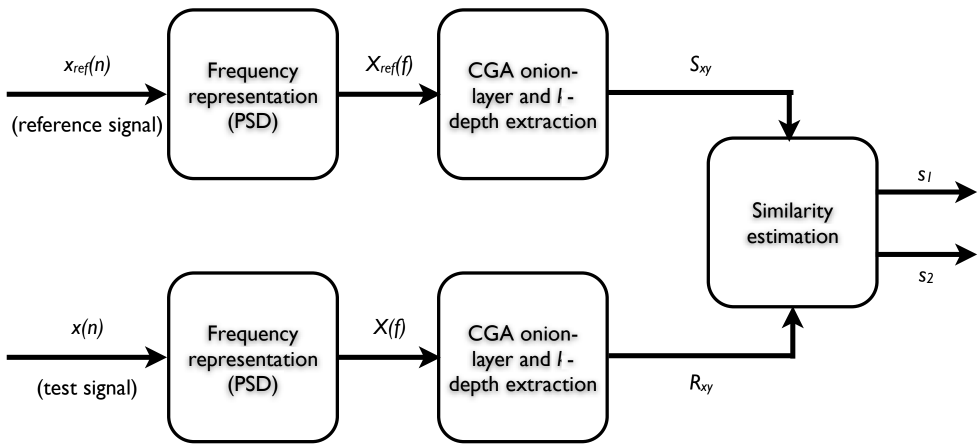


Figure 2. Schematic representation of the preprocessing, feature extraction and identification stages

3. Implementation Issues-Decision Stages

The degrees of correlation s_1 and s_2 (see Equations 1, 2) between S_{xy} and N_{xy} (see Section 2) are calculated according to the selected null hypothesis. The null hypothesis claims that there is no link between the two sampled subsets. Since the distribution of the subsets is unknown, a reasonable strategy is to use a non-parametric approach for testing the hypothesis and thus to use permutations to obtain the subsets distribution under $H_0=0$ with p=0.05 in which all the subsets present random distribution. However, in our case, we used an alternative hypothesis, H_1, which controls the specific similarities between the groups. More specifically, under the current study, we investigated the following three decision stages: (a) the pairs of audio fingerprints are identified; (b) the pairs of audio fingerprints have common features; and (c) the pairs of audio fingerprints are not identified.

For this, the decision rules are extracted by the unsupervised clustering *k*-mean procedure. Thus, one particular issue with the dataset of vector S_i (see Equation 1) is to define the number of categories. In order to achieve a safe result, we use the well-established *k*-means clustering algorithm to examine whether the three targeted decision groups are the optimal clustering approach for the dataset presented later in Table 1. The procedure follows a trivial way to classify a given data set through a certain number of clusters (assuming *k* clusters) fixed *a priori*. The main idea is to define *k* centroids, one for each cluster. These centroids should be placed in a cunning way, because different locations cause different results and a loop evaluates this case recurrently. In our case we submitted the data set (of Table 1) into *k* = 3 clusters, using Equation 3:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| S_i^{(j)} - c_j \right\| \tag{3}$$

where the value $\left\|S_i^{(j)} - c_j\right\|$ represents the selected distance measure between a data point $\vec{S}_i$ and the cluster centre c_j.

The results of the clustering algorithm for the case $k = 3$ are generated from a k-mean processing of 42 vectors that correspond to the audio data presented later on Tables 1-4. These results are graphically illustrated in Figure 3. Having obtained the classification of the vectors S_i (for $k = 3$), additional tests are performed in order to verify that the intersection information is consistent, thus producing robust results not affected by a possibly similar geometrical intersection.

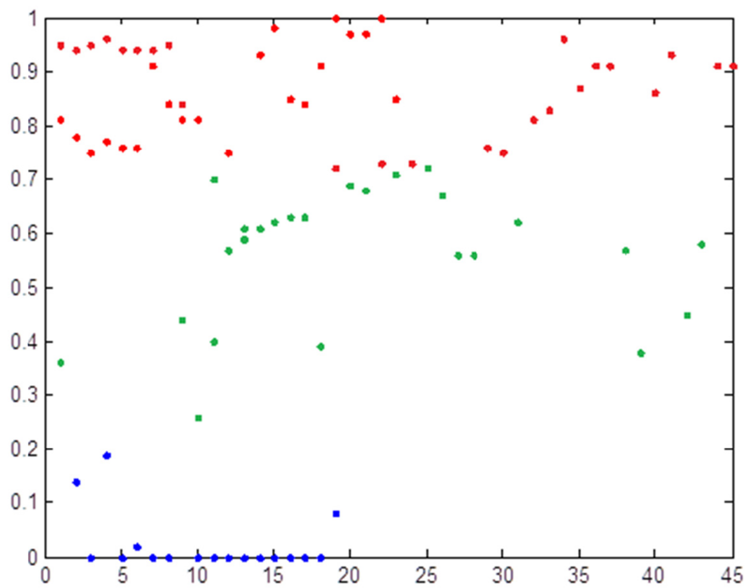


Figure 3. *k*-means clustering procedure. The colored points in blue (non-identified), red (identified) and green (common features) represent all the clustering cases for $k = 3$

Finally, following the above clustering outcome, the decision - threshold values for this work are assigned to three disjoint and exhaustive categories depicted in Table 6 (k means clustering), numerically defined as:

1. $s_1 \geq 0.73$ or $s_2 \geq 0.73$: the compared sampled subsets S_{xy} and N_{xy} are identified.
2. $0.25 \geq s_1 < 0.73$ & $s_2 < 0.73$ or $0.25 \geq s_2 < 0.73$ & $s_1 < 0.73$: the compared sampled subsets S_{xy} and N_{xy} have common features; and
3. $0 \geq s_1 < 0.25$ & $0 \geq s_2 < 0.25$: the compared sampled subsets S_{xy} and N_{xy} are not identified.

4. Results

In this work, we evaluated the proposed audio fingerprinting and classification algorithm efficiency under three different, but typical and widely-employed application scenarios:

a) the compressed application scenario, which aims to assess the identification efficiency of the CGA algorithm for compressed quality audio.

b) the partially-available application scenario, were only a small part of the audio track under identification is available. This represent a typical-everyday case usually supported by music-content retrieval systems.

c) the different performance scenario, were we aim at investigating the identification capabilities of the CGA algorithm on the same audio/music material, performed under different conditions (i.e. recorded in studio or during a live performance).

In all application scenarios, the audio data fed to the CGA algorithm are formatted using the well-known PCM coding scheme (Xing-ji, 2005). The original audio content was encoded using 16bit quantisation resolution and a 44.1kHz sampling rate, following the so-called Compact-disk (CD) sound quality format. Moreover, we particularly employed three different music tracks, with each one belonging to different music genres: a violin solo performance (violin), a rock-band recording (rock) and a symphonic orchestra recording (orchestra). For the purposes of the above application scenarios, this content was compressed using the MPEG-1 Layer III standard at different typical coding bitrates (i.e. 32, 48, 64, 96, 112,128, 160, 192, 256 and 320 kbps) in order to produce the audio content under identification for the previously described application scenarios.

In order to realize the above application scenarios in practice, for the experiments followed, we divided the audio tracks database into two sets:

- The first set is formed by 10 (different typical coding compression schemes) x 3 (different music genres considered) = 30 (total music tracks).
- The second set consists of short-length audio clips (with time-duration equal to 3 seconds). These clips were produced by randomly setting a 3-seconds time window on the compressed audio data. For each audio-recording genre considered (violin, rock and orchestra) we created 5 randomly selected short-audio clips, resulting into a total of 5x30=150 clips belonging into the second dataset.

In the testing procedure, we create 4500 pairs combinations between two sets m=30 and k=150, extracted by Equation 4:

$$Testing\_Number = m * k = 4500 \tag{4}$$

The three investigation scenarios presented earlier allow a number of possible application suggestions to be made. These include copyright applications that include the interactive and fairer calculation of royalty charges, which may then be attributed directly to the copyright owners. Creating a sensing networked device that is installed in order to constantly monitor, identify and report on the commercial use of audio (radio & TV stations) is certainly a novel application which will allow producers the flexibility to interactively select and broadcast content. Also the partial-detection feature may enable charging in a per-second tariff, instead of a flat charge that applies today. The limited sample duration required for this algorithm to detect the source is clearly an advantage that may also boost the commercial application of the method for personal use, enabling the optimisation of current algorithms used in proprietary applications. Finally, the availability of a tool that detects similarity across different versions of the same theme allows a number of tools that may be used as a research or commercial tool to detect audio influences, automatic categorisation of content and copyright issues which could have been missed in the past.

As it will be shown in the next Sections, exhaustive testing across all datasets described above showed that the CGA-based fingerprint-matching algorithm performs well across different compression schemes and most importantly when random segmented tracks are considered. We must also note here that an audio signal time-length equal to 3 seconds outperforms most industrial-level algorithms utilised today in audio proprietary recognition applications which work efficiently for samples greater than 10 seconds (Chandrasekhar et al., 2011).

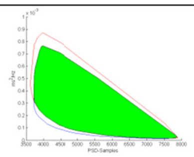
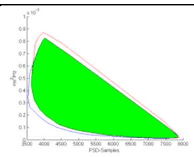
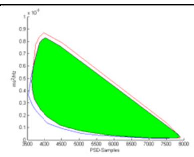
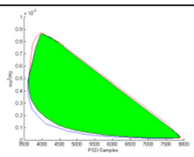
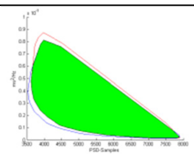
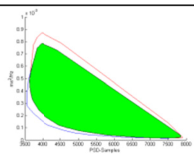
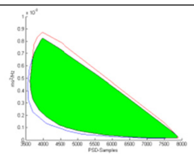
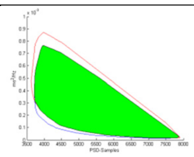
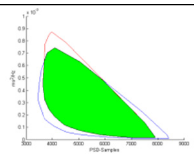
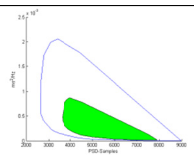
The sequence of tests performed was organised in the following manner: For every music theme we assessed the performance of the proposed algorithm for all considered compression rates, a fact that allowed to identify how well the algorithm performed in music theme recognition at varying compression rates and audio quality. In the testing procedure (Note 1), we considered 4500 pairs of combinations between the two databases sets defined previously.

4.1 Case 1: Compressed Application Scenario

As mentioned previously, under this application case, the three different audio tracks considered as testing material were encoded using the the MPEG-1 Layer III lossy compression standard at different typical coding bitrates, specifically equal to 32, 48, 64, 96, 112,128, 160, 192, 256 and 320 kbps. The compressed content was then decoded, producing a distorted, uncompressed version of the original track. We then applied the CGA algorithm on the original data, as well as on this uncompressed version, and we calculated the degrees of correlation for these signals. Based on the implementation analysis provided in Section 3, we obtained identification or no-identification results for all the combinations of the compressed audio material considered.

A representative set of the above results is presented in Tables 1 and 2. For illustration purposes, we present only the diagrams and values obtained for the violin recordings. Clearly, the CGA algorithm successfully identifies the compressed audio signal correctly, under any selected compression bit rate (even for those that imply lower sampling rates, such as the 32kbps). The same trends were observed for both the rest music tracks considered here (rock and orchestra). This fact is illustrated in Figure 4, were the degrees of correlation s_1 and s_2 values are graphically presented as a function of the compression bit rate. Obviously, in all test cases, at least one of the s_1 or s_2 values exceeds the thresholds defined in Section 3 (and presented in this Figure using the black dashed line), producing a secure and accurate identification outcome.

Table 1. Violin sample track – results and graphical CGA representation

320kbps	256kbps	192kbps
		
$s_1 = 95\%$, $s_2 = 85\%$	$s_1 = 94\%$, $s_2 = 91\%$	$s_1 = 95\%$, $s_2 = 93\%$
Identified	Identified	Identified
160kbps	**128kbps**	**112kbps**
		
$s_1 = 96\%$, $s_2 = 97\%$	$s_1 = 94\%$, $s_2 = 93\%$	$s_1 = 94\%$, $s_2 = 86\%$
Identified	Identified	Identified
96kbps	**64kbps**	**48kbps**
		
$s_1 = 94\%$, $s_2 = 92\%$	$s_1 = 95\%$, $s_2 = 85\%$	$s_1 = 84\%$, $s_2 = 92\%$
Identified	Identified	Identified
32kbps		
		
$s_1 = 26\%$, $s_2 = 100\%$		
Identified		

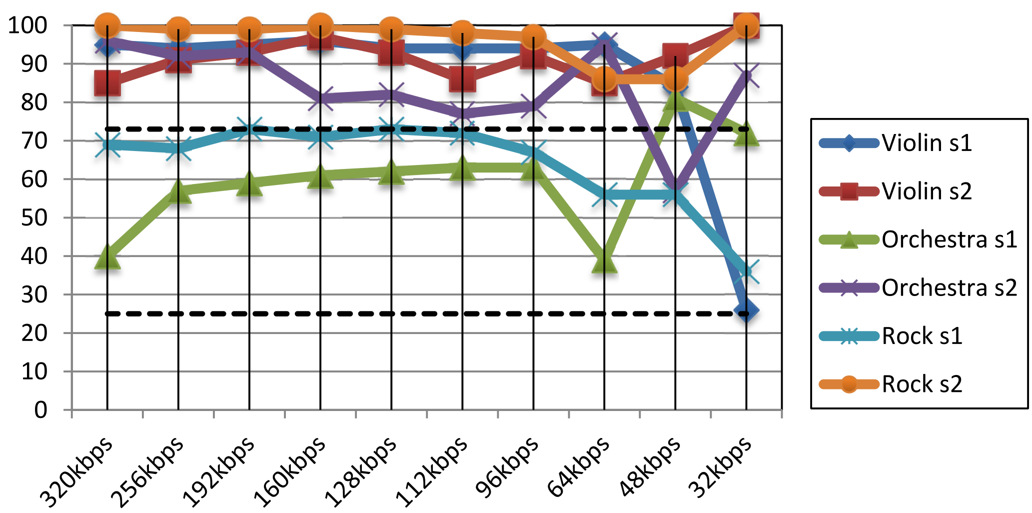
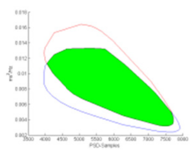
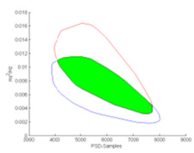
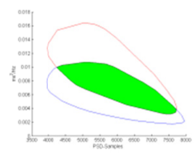
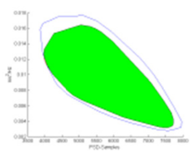
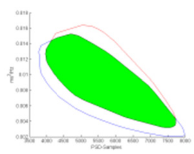
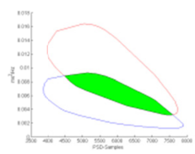
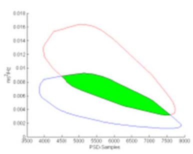
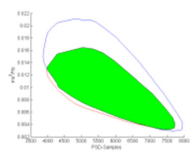
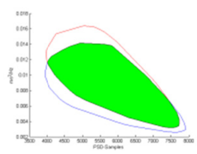
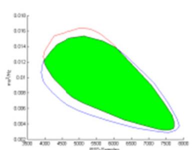


Figure 4. The complete set of the results obtained for the compressed application scenario

4.2 Case 2: The Partially-Available Application Scenario

As mentioned previously, in this application scenario we considered short-length clips of the original audio / music content, produced be randomly selecting a 3 seconds time-length of the compressed and decoded audio material. This case is very typical in everyday life identification and information retrieval applications, where usually, only a short-block of the original music data is available. Table 2 shows the results obtained for the rock-recording test case and for all compression rates considered. Clearly, the CGA algorithm achieves excellent identification rates that are independent of a) the particular audio clip length (since these clips are randomly formed as explained previously) and b) the compression ratio. The same trends can be observed for both the violin and orchestra cases (as illustrated in Figure 5). The only exception in this general trend is the orchestral music track, which, for a compression rate equal to 112kbps, was not clearly identified. We believe that this is attributed to the randomally-selected segment used in the comparison, which was not particularly representative for an orchestral recording with significant dynamic range variations over time. The CGA algorithm in this particular case provided indications of common features between the original and the audio clip under identification. This however, can be considered as a sporadic outcome; thus, it does not impact the overall measured efficiency of the CGA algorithm.

Table 2. Partially-available rock sample track – results and graphical CGA representation

320kbps	256kbps	192kbps
		
$s_1 = 76\%, s_2 = 74\%$	$s_1 = 78\%, s_2 = 50\%$	$s_1 = 75\%, s_2 = 45\%$
Identified	Identified	Identified
160kbps	**128kbps**	**112kbps**
		
$s_1 = 77\%, s_2 = 99\%$	$s_1 = 75\%, s_2 = 82\%$	$s_1 = 76\%, s_2 = 26\%$
Identified	Identified	Identified
96kbps	**64kbps**	**48kbps**
		
$s_1 = 76\%, s_2 = 26\%$	$s_1 = 62\%, s_2 = 94\%$	$s_1 = 81\%, s_2 = 82\%$
Identified	Identified	Identified
32kbps		
		
$s_1 = 83\%, s_2 = 93\%$		
Identified		

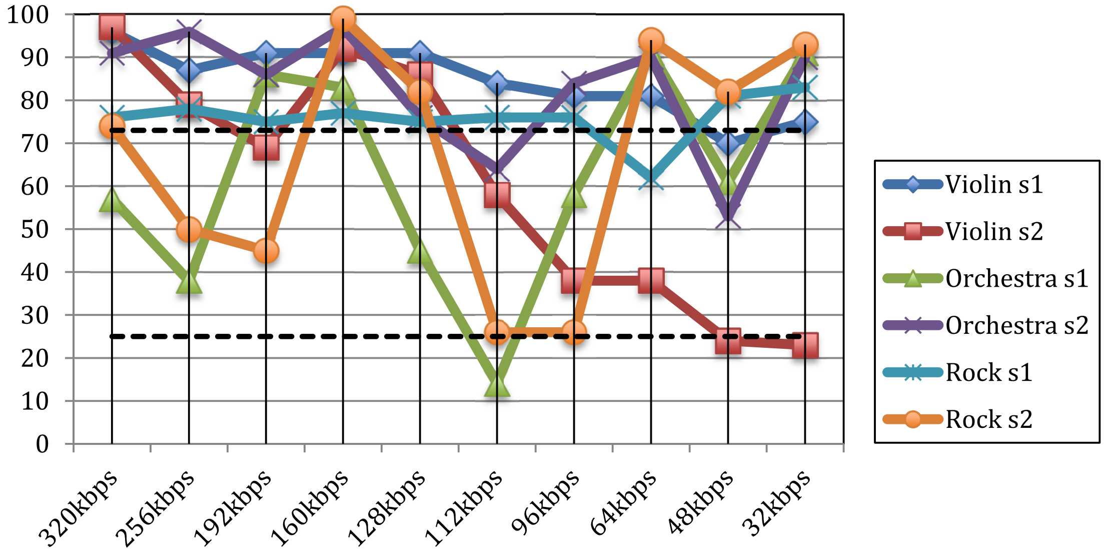
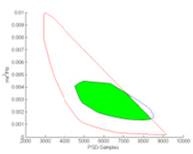
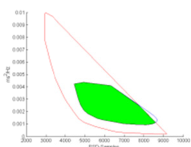
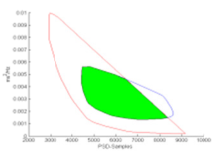
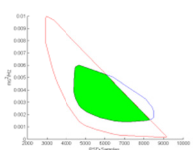
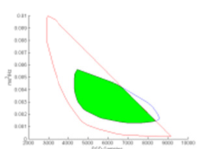
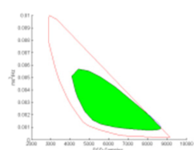


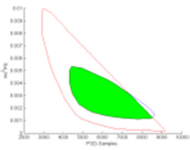
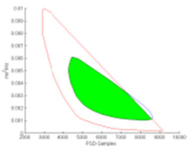
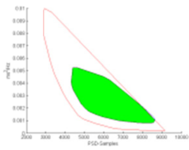
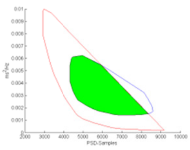
Figure 5. The complete set of the results obtained for the partially-available application scenario

4.3 Case 3: The Different Performance Scenario

In the latter application case considered in this work we investigated the ability and evaluated the performance of the CGA algorithm to identify music material performed under different conditions. In particular, as it was previously mentioned, we employed two original recordings of the same music track, recorded in studio and during a live performance. The specific audio track was a rock one, different from the one employed in the previous tests. We then compressed and uncompressed the live recording under all specified MPEG1-Layer III bit rates, and created corresponding short-length audio clips (again with duration equal to 3 seconds). Finally, the CGA-based identification process was applied, producing the results illustrated in Table 3. Clearly, the identification process efficiency is not affected by the fact that the sound-tracks are recorded in different conditions, while, additionally, it is clearly still independent of the applied compression bit-rate. Table 3 presents varying results and Table 4 contrasts non-identification cases, for clearly diverse data sets such as violin – rock.

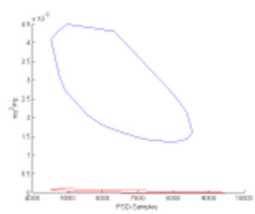
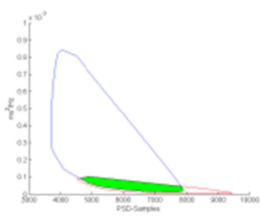
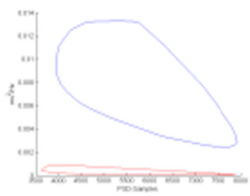
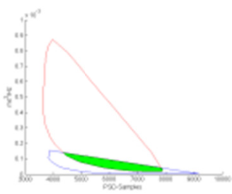
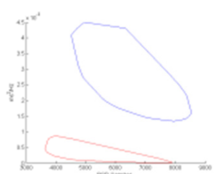
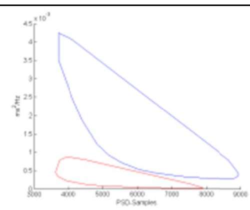
Table 3. Different performance sample tracks – results and graphical CGA representation

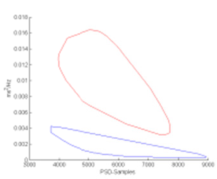
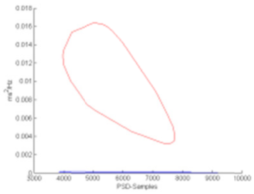
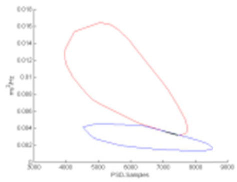
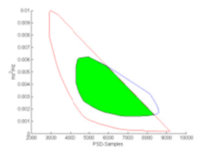
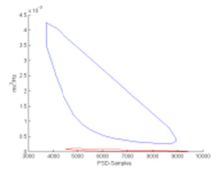
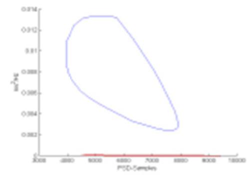
320kbps	256kbps	192kbps
		
$s_1 = 93\%$, $s_2 = 29\%$	$s_1 = 98\%$, $s_2 = 35\%$	$s_1 = 85\%$, $s_2 = 40\%$
Identified	Identified	Identified
160kbps	128kbps	112kbps
		
$s_1 = 84\%$, $s_2 = 42\%$	$s_1 = 91\%$, $s_2 = 41\%$	$s_1 = 100\%$, $s_2 = 48\%$
Identified	Identified	Identified

96kbps	64kbps	48kbps
		
$s_1 = 97\%$, $s_2 = 40\%$	$s_1 = 97\%$, $s_2 = 43\%$	$s_1 = 100\%$, $s_2 = 42\%$
Identified	Identified	Identified
32kbps		
		
$s_1 = 85\%$, $s_2 = 45\%$		
Identified		

For reasons of the results' presentation integrity, in Table 4, additional identification results are hereby presented, by considering the studio recording and the violin, rock and orchestral recordings (all of them in their original formats, i.e. no compression / decompression is employed). Obviously, the CGA-algorithm does not provide any matching results between the different audio tracks. However, it is clearly observed that the algorithm provides an indication of common violin features on a number of contrasted content types. This observation is significant, since it provides a preliminary ability of the CGA algorithm to determine different music genres, a fact that should be investigated in detail in a future work. The same outcome was derived when compression and / or short-length audio-clips were employed instead of the complete-size audio tracks, rendering the CGA fingerprinting algorithm a robust alternative towards compression-, length- and performance-independent music content identification.

Table 4. Results obtained using different types of audio content (no-compression applied)

Orchestra/Concerto	Orchestra/Violin	Violin/Rock
		
$s_1 = 0\%$, $s_2 = 0\%$	$s_1 = 8\%$, $s_2 = 67\%$	$s_1 = 0\%$, $s_2 = 0\%$
Not-Identified	Common Features	Not-Identified
Violin/Orchestra	Violin/Concerto	Violin/Studio
		
$s_1 = 44\%$, $s_2 = 8\%$	$s_1 = 0\%$, $s_2 = 0\%$	$s_1 = 0\%$, $s_2 = 0\%$
Common Features	Not-Identified	Not-Identified

Rock/Studio	Rock/Orchestra	Rock/Concerto
		
$s_1 = 0\%$, $s_2 = 0\%$	$s_1 = 0\%$, $s_2 = 0\%$	$s_1 = 0\%$, $s_2 = 0\%$
Not-Identified	Not-Identified	Not-Identified
Rock/Violin	**Orchestra/Studio**	**Orchestra/Rock**
		
$s_1 = 90\%$, $s_2 = 50\%$	$s_1 = 0\%$, $s_2 = 0\%$	$s_1 = 0\%$, $s_2 = 0\%$
Common Features	Not-Identified	Not-Identified

4.4 Results Aggregation-Statistical Evaluation

A summary of the all the obtained results for all application scenarios when the full dataset is considered (available online, see note at the end of the paper) is presented in Table 5. Clearly, the proposed method achieves very-high scores in terms of correct identification among different audio content types and coding. Also, we executed a Wilcoxon test (Wilcoxon, 1945). We adopted this test because it is possible to make comparisons between two groups using means in paired samples and chi-square analysis. This method is considered more powerful than other non-parametric test paired samples (Little & Rubin, 1987). For the above reasons, we performed three Wilcoxon tests for each category (fully-identified pairs, audio tracks with common features and non-identified content) in order to evaluate the observed differences between them. These differences are analytically illustrated in Table 5.

Table 5. Summary of the identification results

Music Content Pairs	Identified	Non-Identified	Have Common Features	False Identificat ion	True Identifivat ion
Full Quality / Same Tracks Compressed	99	94	5	5	94
Full Quality / Diverse Tracks	0	2092	2309	0	4401
Total	99	**2186**	**2314**	5	4495
Total Experiment		**4500**			

In our case, the null hypothesis is based on the hypothesis test, performed at the 0.05 significance level, in H. H = = 0 indicates that the null hypothesis ("medians are equal or are randomly related"). In Table 6, the Wilcoxon test shows that grouping of the paired data is strongly related. Under *k*-mean clustering the pairs valued (s_1, s_2)=(0, 0) indicate that the null hypothesis is rejexted H = = 1, in which the compared pairs are not correlated. This is the same set depicted in Figure 4 using blue colour.

Table 6. Wilcoxon Paired - Sample test on k-means clustering

S vector		*k* means clustering	Wilcoxon Paired-sample test
s_1	s_2		
0	0	1	p = 0.0020 h =1 The test rejects the null hypothesis of equal medians at the default 5% significance level
0	0	1	
0	0	1	
0	0	1	
0	0	1	
0	0	1	
0	0	1	
0.0200	0.0800	1	
0.0800	0.1000	1	
0.1400	0.1000	1	
0.1900	0.2000	2	p = 0.0036 h =1 The test rejects the null hypothesis of equal medians at the default 5% significance level
0.2600	0.2100	2	
0.5600	0.3800	2	
0.5800	0.4100	2	
0.5900	0.4200	2	
0.6100	0.4200	2	
0.6100	0.4300	2	
0.6200	0.4500	2	
0.6300	0.4800	2	
0.6300	0.5000	2	
0.7200	0.7400	3	p = 0.0405 h =1 The test rejects the null hypothesis of equal medians at the default 5% significance level
0.7300	0.7700	3	
0.7500	0.7900	3	
0.7600	0.8200	3	
0.8100	0.8600	3	
0.8100	0.8600	3	
0.8400	0.9000	3	
0.8400	0.9000	3	
0.8500	0.9100	3	
0.8500	0.9100	3	
0.8600	0.9200	3	
0.9100	0.9200	3	
0.9100	0.9200	3	
0.9100	0.9300	3	
0.9100	0.9300	3	
0.9400	0.9700	3	
0.9600	0.9800	3	
0.9600	0.9900	3	
0.9700	0.9900	3	
0.9700	0.9900	3	
0.9800	1.0000	3	
1.0000	1.0000	3	

5. Conclusions

This study's major objective was to try to evaluate a mechanism of audio detection similarity and fingerprinting that, under certain circumstances, can be inserted into the information-management strategies of (large) information organisations and large, co-operative libraries as a technique for the identification and possible control of the intellectual property of electronically published audio material. The existing techniques, and especially the digital signature schemes, could fulfill only the first, the identification and part of the objective.

In particular, this work employs the adjustment of a computational geometric algorithm for the semantic representation of the information of audio data in terms of a frequency-domain audio fingerprint. The idea for this construction came from the test of the onion-peeling algorithm in other areas of signal processing, such as the identification of humans by fingerprints. The aim of this application is to construct an audio fingerprint (i.e. in terms of a serial number) that could identify a copyright-protected published audio file even if its file format has changed from one type to another. Furthermore, it aims to provide a satisfactory amount of correlation similarity with other audio files created from the original by applying different coding / compression techniques, and to detect and automatically reject audio files that are not related to the original.

For a realistic implementation and efficiency assessment of the proposed audio fingerprinting algorithm, the authors created a small database with three different audio genres encoded using the MPEG-1 Layer III specification at multiple compression ratios, enabling experimentation with internal and external data-sets. This demonstrated the computational efficiency of the algorithm, which was sucessfully used under three different application scenarios: the first investigates matching of a full audio clip duration using varying compression settings. In the second scenario compression and sample duration vary while the third introduces context testing across different performances and orchestrations. This last scenario introduces a "common feature" tracking mechanism, which allows automated comparison of different audio tracks that share musicological characteristics. For instance, we found that the audio clip containing the violin may be partially associated with instrumental tracks containing the violin (orchestra file), a characteristic that was consistent across a wide variety of experimental executions.

We additionally proved, via the Wilcoxon sample paired test, that the categories of the intersection areas between different or same audio clips are related strongly. We also found that the fingerprint features must be aligned temporally; that is, if a set of features appears in both the original recording in the database and in a sample query, the relative positions of each feature within each recording must be the same. The computational load of the algorithm behaves linearly (i.e. *O(2n)*) for each comparing tuple and may be bounded with a second-order polynomial for the comparison procedure *O(2n^2)* under the worst-case scenario.

Our future research on this work topic will focus on the comparison of the algorithmic results for data with varying similarity. Mixed audio tracks and segments may be used for pattern matching, enabling automated copyright-verification to be performed. The authors believe that the same algorithm may also be utilised for other multimedia data types including images, video, text and combined applications such as web pages, multimedia systems and databases.

Acknowledgements

We would like to acknowledge the assistance of Nicos Skiadopoulos for the development of the web-application, enabling researchers to dynamically test our data or their own.

References

Aucouturier, J. J., & Pachet, F. (2003). *Music similarity measures: Whats the use?* Paper presented at the International Symposium on Music Information Retrieval.

Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., & Slaney, M. (2008). Content-based music information retrieval: current directions and future challenges. *Proceedings of the IEEE, 96*(4), 668-696. http://dx.doi.org/10.1109/JPROC.2008.916370

Chandrasekhar, V., Sharifi, M., & Ross, D. A. (2011). *Survey and Evaluation of Audio Fingerprinting Schemes for Mobile Query-By-Example Applications*. Paper presented at the International Conference on Music Information Retrieval (ISMIR).

Deliyannis, I. (2012). From Interactive to Experimental Multimedia. In I. Deliyannis (Ed.), *Interactive Multimedia* (pp. 3-12). Rijeka, Croatia: Intech.

Deliyannis, I., Karydis, I., & Anagnostou, K. (2011). *Enabling Social Software-Based Musical Content for Computer Games and Virtual Worlds*. Paper presented at the 4th International Conference on Internet

Technologies and Applications (ITA2011).

Deliyannis, I., Karydis, I., & Karydi, D. (2011). *iMediaTV: Open and Interactive Access for Live Performances and Installation Art*. Paper presented at the 4th International Conference on Information Law (ICIL2011).

Furht, B., & Kirovski, D. (2005). *Multimedia Security Handbook* (illustrated ed.). CRC Press.

Gillespie, T. (2004). Copyright and Commerce: The DMCA, Trusted Systems, and the Stabilization of Distribution. *The Information Society Journal, 20*, 239-254. http://dx.doi.org/10.1080/01972240490480938

Karydi, D., Karydis, I., & Deliyannis, I. (2012). *Legal Issues in Using Musical Content from iTunes and YouTube for Music Information Retrieval*. Paper presented at the International Conference on Information Law.

Karydis, I., Deliyannis, I., & Floros, A. (2011). *Augmenting Virtual-Reality Environments with Social-Signal Based Music Content*. Paper presented at the 17th International Conference on Digital Signal Processing (DSP2011). http://dx.doi.org/10.1109/ICDSP.2011.6004944

Kosch, H. (2004). *Distributed multimedia database technologies: supported by MPEG-7 and MPEG-21*: CRC Press.

Levy, M., & Sandler, M. (2009). Music information retrieval using social tags and audio. *IEEE Transactions on Multimedia, 11*(3), 383-395. http://dx.doi.org/10.1109/TMM.2009.2012913

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data* (Vol. 2). New York: Wiley.

Logan, B., Ellis, D. P. W., & Berenzweig, A. (2003). *Toward evaluation techniques for music similarity*. Paper presented at the SIGIR 2003: Workshop on the Evaluation of Music Information Retrieval Systems.

Marsden, A. (2010). *Recognition of Variations Using Automatic Schenkerian Reduction*. Paper presented at the International Society for Music Information Retrieval.

McFee, B., Barrington, L., & Lanckriet, G. (2010). *Learning similarity from collaborative filters*. Paper presented at the International Society for Music Information Retrieval.

McKay, C., & Fujinaga, I. (2008). *Combining features extracted from audio, symbolic and cultural sources.* Paper presented at the International Conference on Music Information Retrieval.

Poulos, M., Deliyannis, I., & Floros, A. (2012). Audio Fingerprint Extraction using an Adapted Computational Geometry Algorithm. *Computer and Information Science, 5*(6), 88-97. http://dx.doi.org/10.5539/cis.v5n6p88

Simpson, S. (2004). Explaining the commercialization of the internet: A neo-Gramscian contribution. *Information, Communication & Society Journal, 7*(1), 50-68. http://dx.doi.org/10.1080/1369118042000208898

Slaney, M., Weinberger, K., & White, W. (2008). *Learning a metric for music similarity*. Paper presented at the Information Society for Music Information Retrieval.

Trifonova, A., Jaccheri, L., & Bergaust, K. (2008). Software engineering issues in interactive installation art. *Int. J. Arts and Technology, 1*(1), 43-65. http://dx.doi.org/10.1504/IJART.2008.019882

Wang, A. (2003). *An industrial strength audio search algorithm*. Paper presented at the International Conference on Music Information Retrieval (ISMIR).

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin, 1*(6), 80-83. http://dx.doi.org/10.2307/3001968

Xing-ji, G. (2005). Structure ansd Application pf WAV File. Retrieved from http://www.elecfans.com/

Note

Note 1. In order to enhance comprehension and allow external verification of the experimental results, the reader may access a web-based application under the following URL:
http://lit.ionio.gr/index.php?option=com_content&view=article&id=70&Itemid=101&lang=el