

Audio Fingerprint Extraction Using an Adapted Computational Geometry Algorithm

Marios Poulos¹, Ioannis Deliyannis² & Andreas Floros²

¹ Laboratory of Information Technologies, Dept. of Archives and Library Science, 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, GR-49100 Corfu, Greece, Canada. Tel: 1-30-26610-87429. E-mail: mpoulos@ionio.gr

Received: July 18, 2012 Accepted: August 15, 2012 Online Published: October 25, 2012

doi:10.5539/cis.v5n6p88

URL: <http://dx.doi.org/10.5539/cis.v5n6p88>

Abstract

This work presents an adapted version of the Computational Geometry Algorithm (CGA) used for the development of audio-based applications and services. The CGA algorithm analyses an audio stream and produces a unique set of points that can be considered to be the audio data “fingerprint”. It is shown that this fingerprint is coding-independent, a fact that can render the proposed algorithm suitable for multiple purposes, 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. Additionally, based on specific novel applications and services, the overall algorithmic performance and efficiency characteristics of the CGA algorithm are discussed and analysed.

Keywords: Computational Geometry Algorithm, fingerprint, multiple purposes, performance, efficiency

1. Introduction

The use of digital technology for content distribution and reproduction introduces new domains and audio-embedded applications. New-media arts (Trifonova, Jaccheri, & Bergaust, 2008), interactive and experimental multimedia (Deliyannis, 2012) are characteristic areas of application featuring dynamic user-content interaction (Deliyannis & Karydis, 2011) and a culture of continuing content use changes (Gillespie, 2004). These global transformations introduce new services and markets (Simpson, 2004), a fact that is supported by the evolution of standards such as MPEG-7, which links content to context offering multimedia accessibility, while MPEG-21 is designed to support full user-content interaction. The considerable content availability through various media certainly induces requirements for new broadcast-control and content verification mechanisms. Additionally, the area of copyright management shares cross-media solutions such as pattern-matching algorithms and techniques to identify copyrighted content in various media types (Furht & Kirovski, 2005). All the above uses require digital coding audio format and sampling ratio – independent identification and fingerprinting, able to cover a wide range of music genres.

Audio fingerprinting techniques aim to develop mechanisms for assessing the perceptual equivalence of different audio / audio content. This is performed by providing a summary of the corresponding audio clip, which is typically stored in a database and serves as an index to the audio library metadata (Beekhof, Voloshynovskiy, Koval, & Holotyak, 2009). By further using matching / similarity functions, distorted versions of the audio content can be mapped to the original version. Particular research issues relating to audio fingerprinting typically include: (i) the determination of the limit that an audio clip can be processed before it cannot be identified anymore (degradation) (Beekhof et al., 2009); (ii) the assessment of the probability of false identification (reliability) (Kennedy & Naaman, 2009; Turnbull, Barrington, Torres, & Lanckriet, 2008); (iii) the effect of the audio segment length on the identification efficiency (granularity) (Marolt, 2008; Orio, 2010; Seifert, 2004); and (iv) the search speed, i.e., the time required for accessing a fingerprint in a large database (Seo et al., 2005).

Due to the several forms of audio content processing and transformation, an efficient audio fingerprinting technique should provide accurate, reliable and robust identification against a large number of audio-processing tasks, including, for example, lossy compression, equalisation, time-scale modification, and linear speed change

(Seo et al., 2005). Towards this aim, different fingerprinting algorithm realisation approaches exist (Baluja & Covell, 2006; Seifert, 2004; Sukittanon & Atlas, 2002). Some works cited in the literature utilise specific peaks extracted from the Power Spectral Density (PSD), which may be visualised through a “constellation map” (Wang, 2003). In particular, they derive PSD peaks and represent them using a scatter plot. This approach imposes a time-alignment problem between the original audio and the comparison audio peaks in the “noisy” (distorted) case (Baluja & Covell, 2006; Chandrasekhar, Sharifi, & Ross, 2011), as it detects a significant cluster of points that form a diagonal line within the scatterplot in order to perform matching using as basis the density criterion (Wang, 2003). In that respect, successful matching is implemented via the identification of a similar diagonal line within the scatterplot of the original file stored in the library. In order to increase accuracy, proprietary mobile phone software applications, such as Shazam (shazam.com) and Sound Hound (soundhound.com), demand additional audio that requires samples of over 10 seconds for matching and 30-60 seconds for copyright detection (Chandrasekhar et al., 2011).

The present work offers a novel approach under the identification (matching) stage, as it is based on the significant reduction of the original spectral peaks enclosed in convex layer areas. The proposed approach introduces audio-track identification through the use of computational geometry algorithms. The problem of matching sample peaks with original peaks is addressed using an intersection technique between convex layers. More specifically, the proposed method produces a convex polygon in the frequency domain that resembles a coordinate-based fingerprint pattern. This is performed through the Computational Geometry Algorithm (CGA), a scheme of onion-like layers that results in unique frequency-domain representations of the innermost onion layer (Poulos, Belesiotis, & Papavlasopoulos, 2007; Poulos, Bokos, & Vaioulis, 2008; Poulos, Magkos, & Chrissicopoulos, 2003; Poulos, Papavlasopoulos, Bokos, Avlonitis, & Kanellopoulos, 2007; Poulos, Papavlasopoulos, Chrissicopoulos, & Magkos, 2003).

The paper is organized as follows. In Section 2, the analytic description of the proposed CGA audio fingerprinting algorithm is presented. Section 3 describes the proposed identification procedure using the onion algorithm in relation to the density criterion that is commonly applied in similar identification problems. Finally, in Section 4 the further research and the concluded remarks are discussed.

2. CGA Algorithm Description

As mentioned in the previous Section, the fundamental requirement for robust and reliable audio fingerprinting depends on the extraction of features that are independent from audio encoding techniques (quality) or common signal degradation. Moreover, the fingerprint produced for each sampled audio clip represents a significant parameter that may affect the realisation complexity, mainly in terms of computational load and memory requirements. Both the above requirements are met by the reduction of the spectral resolution of the original Fast Fourier Transform (FFT) magnitudes using the Onion Algorithm (OA), according to our latest studies (Gillespie, 2004; Kosch, 2004; Trifonova et al., 2008). In more detail, the basis of the method depicts the centre of the multi-layers of a set of arithmetic points on the Cartesian plane that represent the values of the application used (Dalal, 2004). In the case of fingerprint audio verification (Poulos et al., 2007), these layers can be formed by the values of the audio signal FFT magnitudes. Additionally, the basis of this algorithm has been already applied successfully in a text categorisation and audio identification procedures (Poulos et al., 2003, 2007, 2008;). More specifically, the proposed CGA audio fingerprinting technique is divided into three stages, namely pre-processing, feature extraction and identification, which are described in the following Sections.

2.1 Pre-processing Stage

This stage is divided into two stages: audio digitisation and spectral processing.

In **audio digitisation**, the audio signal under identification is processed in order to be transformed / decoded in a pre-defined reference coding scheme. For the purposes of this work, the reference coding-scheme considered was the widely employed linear Pulse Code Modulation (PCM) method. This was the coding scheme of all the reference (initial) signals employed in this work for assessing the proposed audio fingerprinting algorithm efficiency. Moreover, the pre-processing stage is also responsible for removing any additional data produced by the initial audio coding process. For example, assuming that the audio signal under identification is compressed according to the MPEG-1 Layer III standard, the initial fixed-duration silence imposed by the relative coder (if any) is removed.

In **spectral processing**, the Power Spectral Density (PSD) P_{xx} of the pre-processed signal under identification x is calculated in units of power per radians per sample. The corresponding vector of frequencies w is computed in radians per sample, and has the same length as P_{xx} . It should be noted that a real-valued input vector x produces a full power one-sided (in frequency) PSD (by default), while a complex-valued x produces a two-sided PSD. In

general, the length N of the FFT and the values of the input x determine the length of Pxx and the range of the corresponding normalised frequencies. For this syntax, the (default) length N of the FFT is the larger of 256 and the next power of 2 greater than the length of x .

Consider that the pre-processed sample $\{x_t\}_{t=0}^{N-1}$ of a stationary, discrete-time, real-valued signal with power spectral density, or spectrum, $\Phi(\omega)$ ($\omega \in (-\pi, \pi]$) For simplification we considered that $\Phi(p)$ denotes the values of the spectrum at the Fourier frequency grid points:

$$\omega_p = \frac{2\pi}{N} p, p=0, \dots, N-1 \quad (1)$$

The periodogram estimate of $\Phi(p)$ is given by (Stoica & Moses, 1997) as:

$$\hat{\Phi}_p = \frac{1}{N} \left| \sum_{t=0}^{N-1} x(t) e^{-i\omega_p t} \right|^2, p=0, \dots, N-1 \quad (2)$$

Thus, the periodogram values consist of the elements of the vector S :

$$\mathbf{S} = \begin{bmatrix} |\Phi_1| \\ \cdot \\ \cdot \\ \cdot \\ |\Phi_p| \end{bmatrix} \quad (3)$$

2.2 Feature Extraction Stage

In the feature extraction stage, the OA method is based on a novel statistical estimation in which the smallest layer of an onion convex polygon encloses the geometric median value of a feature vector (Poulos et al., 2007). Additionally, the semantic features of the OA are grounded in previous studies, in which, the smallest convex polygon (layer) depicts a particular geometrical area in which this average value of the data may be characterised as strong effect on the properties of the data (Dalal, 2004) and unique (Poulos et al., 2003). The OA has also been widely used in many application domains, i.e., image processing, pattern recognition, photo image analysis, and study of Earth atmosphere (Bae, Alkobaisi, Vojtechovsky, Narayanappa, & Bae, 2010) in which the process of peeling a planar point set is central in the study of robust estimators in statistics (Chazelle, 1985). The feature extraction of this study is based on a procedure, in which the peaks in each time-frequency according to the amplitude are chosen via inner layer (suitable depth) of the peak's amplitudes. The choice of the critical depth of the OA is estimated via the geometric property of convex hull where the onion technique is based upon allows one to scale, rotate, and shift the spatial structure without changing its property (Chang et al., 2000). Thus, in the experimentation we selected the suitable depth's layer (k) (see Equation 4), where the variance of the enclosed amplitudes was minimised. In this way, we applied Bartlett's estimation in spectral resolution for a reduction in variance in order to isolate minimised variance according to Equation (4):

$$F = \text{Var} \left\{ \hat{\Phi}_p(\omega) \right\} \approx \frac{1}{k} \hat{\Phi}_p^2(\omega) \quad (4)$$

Additionally, this statistical approximation has been verified empirically in several pattern recognition problems (Bose & Toussaint, 1995; Graham, 1972; Hess, 1983; O'Rourke, 1998). The proposed procedure is divided into the following 3 steps:

(1) In the Cartesian plane (see Figure 1) the values of vector S are placed in the y -axis and in the x -axis the order of each element of vector S is shown. We put the elements of the S vector in the Cartesian plane according to the $f(p, |\Phi_p|)$ function.

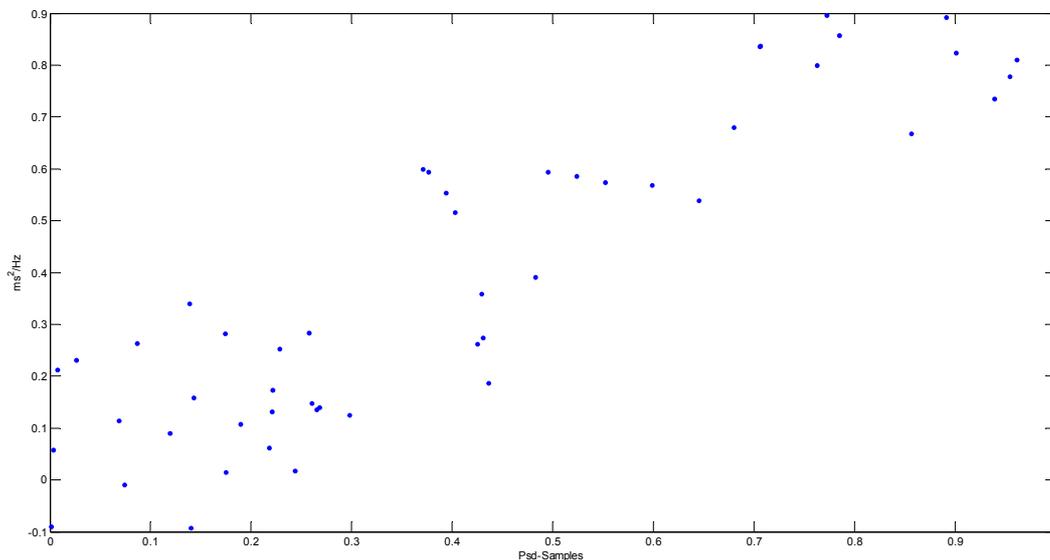


Figure 1. The placement of coordinates $f(p, |\Phi_p|)$

(2) We determine the finite set of points $S=S_0$. Let S_1 be the set $S_0 \setminus \partial H(S_0):S$ minus all the points on the boundary of the hull of S , as it is illustrated in Figure 2:

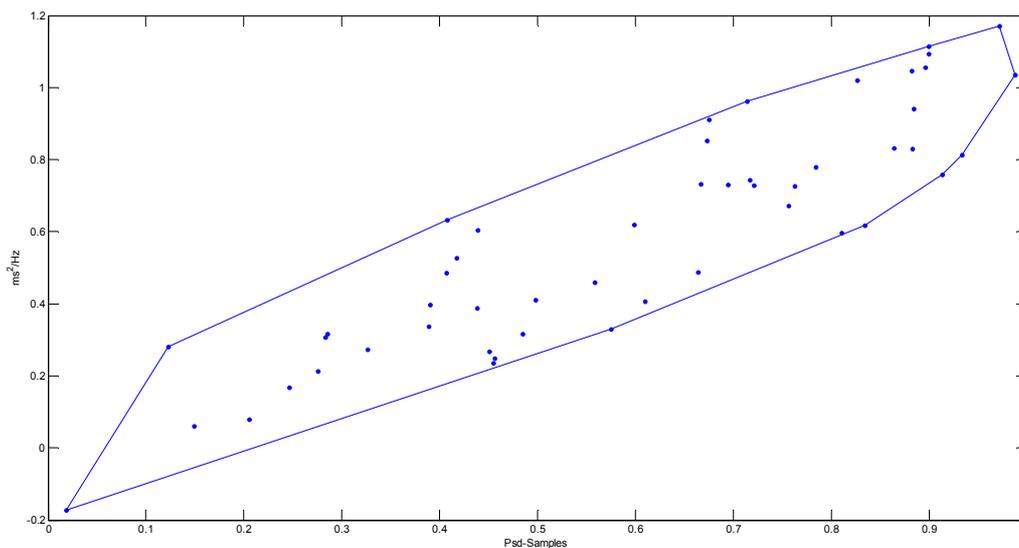


Figure 2. The external hull S_0

(3) The process continues until a set with three innermost points or less is produced. Similarly, we define $S_{i+1} = S_i \setminus \partial H(S_i)$. The Hulls $H_i = \partial H(S_i)$ are also referred as *layers* of the set and the process of peeling away the layers is termed *onion peeling* (Graham, 1972; O'Rourke, 1998) (see Figure 3).

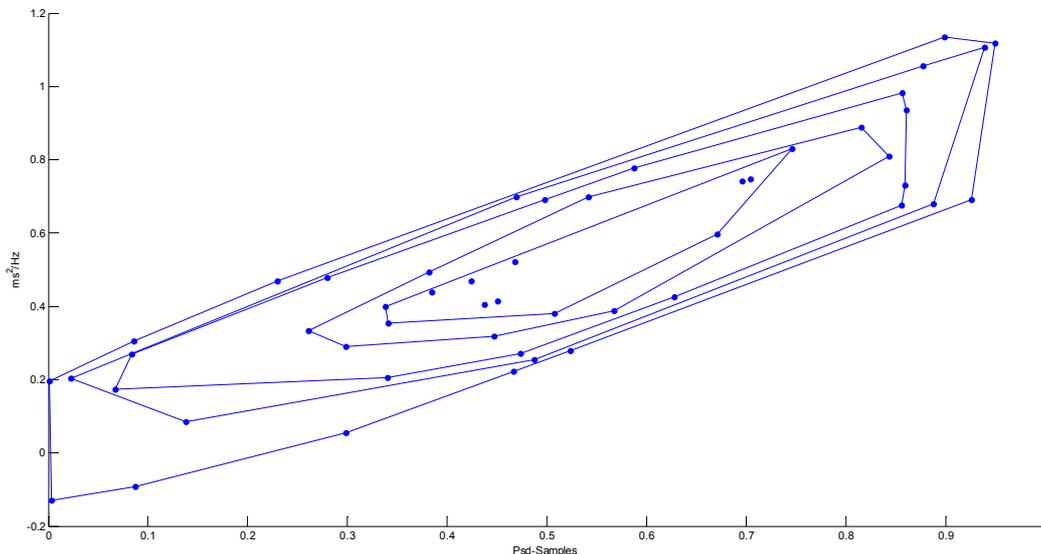


Figure 3. The iterative procedure of convex hulls calculation

This position may be determined by using a combination of computational geometry algorithms, which is known as Onion Peeling Algorithms (O'Rourke, 1998), with overall complexity $O(k*n \log n)$ times, where k (see equation 4) is the depth of the convex layer and n is the size of vector S (see equation 3) in the numerical representation (in accordance with the description of the pre-processing stage in Section 2.1). However, according the equation 4 the calculation, the value k is implemented as follows:

- The standard least-squares technique was selected as opposed to the non-linear minimisation method of Newton - Raphson because it can estimate k more efficiently.
- Convergence of the algorithm described above is not guaranteed, as it was verified in the experimental part. For those $\hat{\Phi}_x^2(\omega)$ amplitudes of the envelope candidate k ultimate layer where convergence is not achieved, we resort to the non-linear minimisation of $\hat{\Phi}_x^2(\omega)$ via the Newton-Raphson method. It has been seen in the experimental part that in almost all such cases, the parameters have converged. In the rare cases where convergence failed to occur, candidate layers were deleted from the set.
- When applying the Newton-Raphson minimisation, the parameter estimate $\hat{\Phi}_x^2(\omega)$ obtained through the least squares minimisation is used as the initial point of the iteration. The first and second derivatives of $\hat{\Phi}_x^2(\omega)$ are estimated and the respective derivative matrices $F'(\hat{\Phi}_x^2(\omega))$ and $F''(\hat{\Phi}_x^2(\omega))$ are formed and used in the iteration:

$$\left(\hat{\Phi}_x^2(\omega)\right)^{k+1} = \left(\hat{\Phi}_x^2(\omega)\right)^k - \left(F'(\hat{\Phi}_x^2(\omega))^k\right)^{-1} F''(\hat{\Phi}_x^2(\omega))^k \tag{5}$$

Here, $\left(\hat{\Phi}_x^2(\omega)\right)^k$ is the estimate produced at the k -th step of the iteration. The convergence threshold, which is the Euclidean distance between every two consecutive estimates $\left\|\left(\hat{\Phi}_x^2(\omega)\right)^{k+1} - \left(\hat{\Phi}_x^2(\omega)\right)^k\right\|_2$, is taken equal to 0.1.

In our example, the calculation of this depth k is valuable because this subset may be characterised as representing the significant semantics of the selected audio signal (see Figure 4). The decision regarding this selection is explained in Section 3.

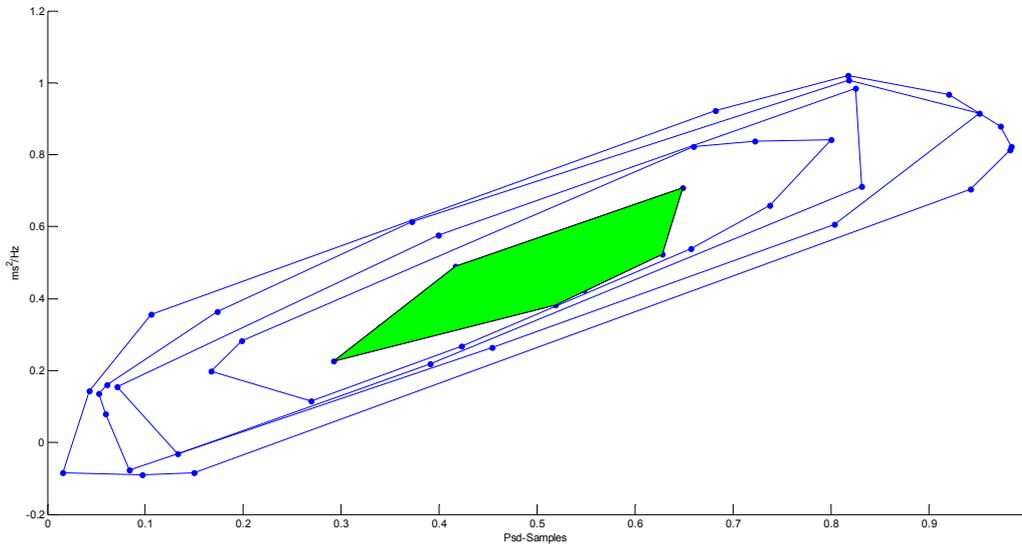


Figure 4. Graphical representation of the ultimate (last) convex polygon, which has depth k

We may consider the smallest convex layer to comprise a significant geometrical region of frequency enclosing the median frequencies of the original audio shot file.

3. Identification Stage

In this stage subset S_{xy} intersects a new subset N_{xy} , which came from the processing of another set N (see Section 2.3), on the Cartesian plane, as illustrated in Figure 5:

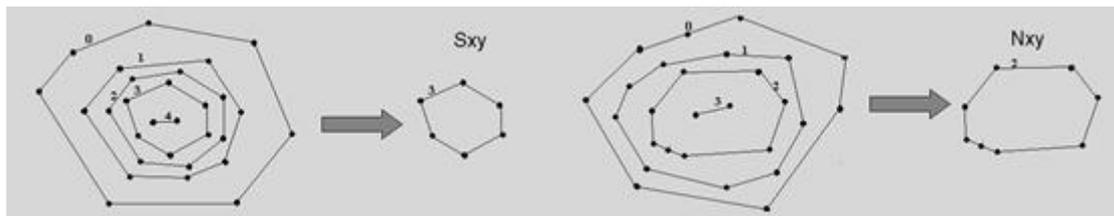


Figure 5. Decision stage between two different onion algorithm procedures

The correlation of investigated subsets S_{xy} , N_{xy} is decided based on the following experimental conditions: if the above subsets are intersected, the degree of correlation is extracted by using the fraction of the common intersection and the area of the original convex set. In more detail, if we consider that $S_{xy} \cap N_{xy} = R_{xy}$ and the corresponding areas of these convex polygons are AS_{xy} , AN_{xy} and AR_{xy} , then the degree α of the correlation of subset S_{xy} with regard to subset N_{xy} is extracted as follows (see Figure 6):

$$s_1 = \frac{AR_{xy}}{AS_{xy}} \tag{6}$$

$$s_2 = \frac{AR_{xy}}{AN_{xy}} \tag{7}$$

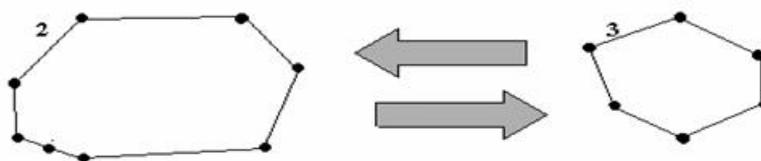


Figure 6. The procedure of correlating the two (2) smallest convex polygons, extracted using the onion layer method

In this way, the degrees of similarity s_1 & s_2 consist of the vector:

$$\vec{S}_i = [s_1, s_2] \tag{8}$$

where $i=1, \dots, x$. More details of the degrees of similarity s_1 & s_2 are given in the Experimental Part depicted in Figure 7.

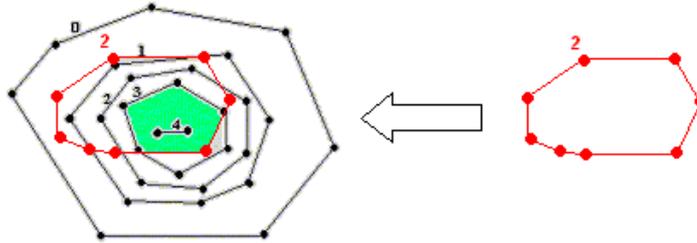


Figure 7. Case of categorisation in the same category. The green area is the common area of the compared text categorisation documents

3.1 Related Study

As mentioned in the introduction, the density criterion is the dominant method in which the known methods are based (Wang, 2003). According to this criterion the candidate peaks are used to assure that the time-frequency strip for the audio file has reasonably uniform coverage. Thereinafter, the selected peaks from the each side are ordered so that diagonal elements correspond to target/target distances. Thus, ideally a threshold θ can be chosen such that all diagonal elements are smaller than, and all off-diagonal elements are larger than, θ (Burges, Platt, & Jana, 2003). For a rectangular window see Figure 8; the transform kernel $\hat{\Phi}_p$ (see equation 2) has a $\sin x/x$ behavior along the heaviest diagonal line (Sukittanon & Atlas, 2002), in which a threshold θ is obtained by the distance of the centroid of p in relation to of diagonal (see Section 2.1). In our case, the threshold is the value of the fraction between the intersected convex layers, which expresses the degree of correlation between two audio files (clips). Thus, the proposed method is superior to known methods because the threshold is satisfied with a non-linear way (Milenkovic, 1999) via degree of intersection (see Section 2.3). In this case, the absolute alignment between the amplitudes of the peaks with a minimum threshold error is not needed (as shown in Figure 10). For better comprehension, the phases of the comparison selected amplitude pairs of the peaks in two algorithms are described in Figures 8-10.

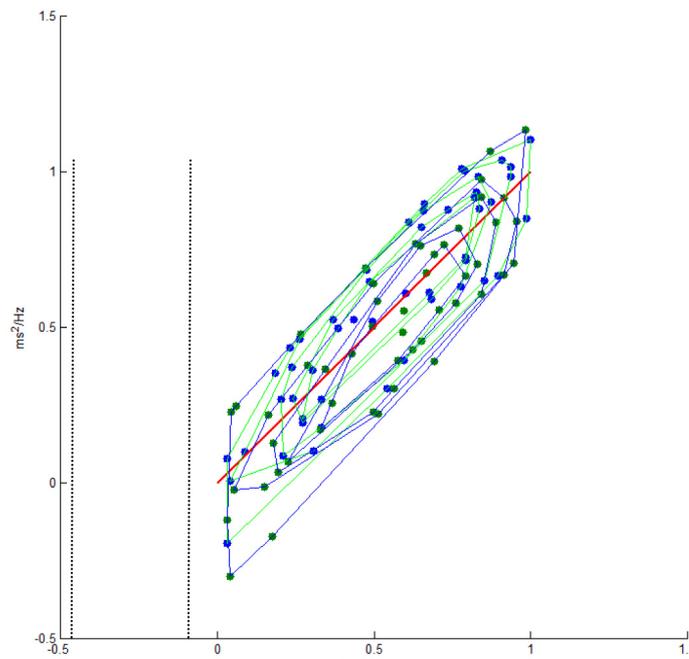


Figure 8. The red line depicts the diagonal elements (peaks) correspond to target/target distances while the onions peeling graphs, with green for the original audio file and blue for the comparison, are also presented

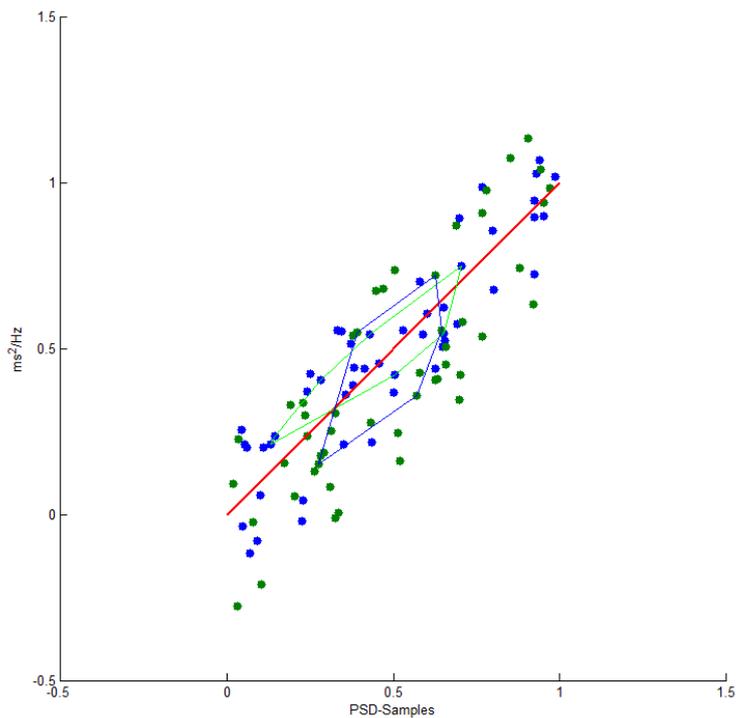


Figure 9. In the second phase, the two ultimate layers (convex polygons) are depicted in relation to with the position of elements' diagonal

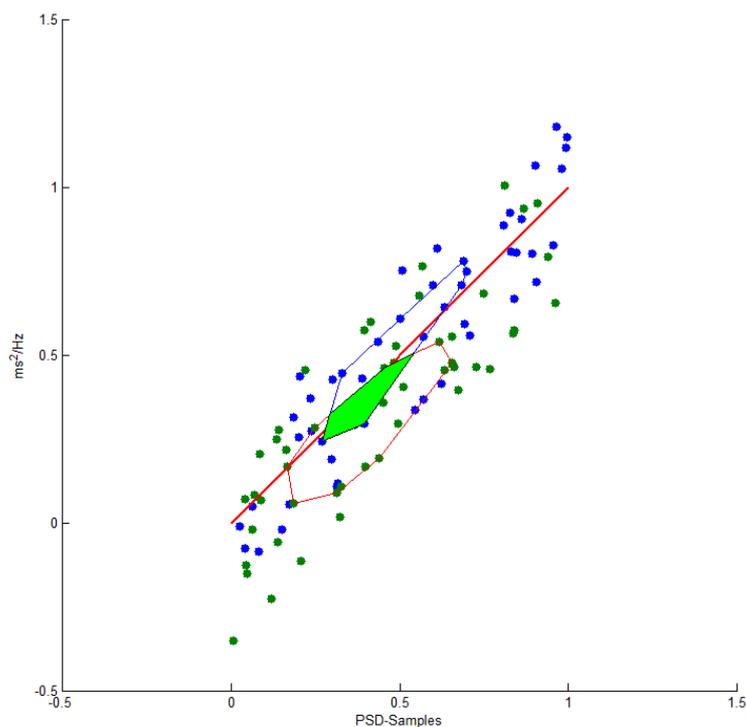


Figure 10. In the third phase, the common section (green color) of the two ultimate layers (convex polygons) is depicted in relation to with the position of elements' diagonal

4. Further Research and Conclusions

The research field addressed in this work is broad and interdisciplinary, incorporating not only aspects of computer science, but in our case diverse fields such as archival science, cognitive science, commerce,

communications, law, library science and signal processing which are essentially interconnected. Therefore, the need for interdisciplinary research in the area of information integration and computer sciences is evident from this work. The experimental aspects of the algorithm have to be addressed in the near future while standardization of the identification process is the key issue that needs to be established, before this technology may be exploited commercially. Our preliminary studies showed that beyond one-to-one pattern matching, dynamic content detection is also possible (Poulos et al., 2007). 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.

This study's major objective was to try to construct an audio mechanism of detection similarity and fingerprinting. This work presents 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.

References

- Bae, W. D., Alkobaisi, S., Vojtechovsky, P., Narayanappa, S., & Bae, K. Y. (2010). *Convex Onion Peeling Genetic Algorithm: An Efficient Solution to Map Labeling of Point-Feature*. Paper presented at the ACM Symposium on Applied Computing.
- Baluja, S., & Covell, M. (2006). *Content fingerprinting using wavelets*. Paper presented at the 3rd European Conference on Visual Media Production CVMP. <http://dx.doi.org/10.1049/cp:20061964>
- Beekhof, F., Voloshynovskiy, S., Koval, O., & Holotyak, T. (2009). *Fast identification algorithms for forensic applications*. Paper presented at the First IEEE International Workshop on Information Forensics and Security, WIFS. <http://dx.doi.org/10.1109/WIFS.2009.5386480>
- Bose, P., & Toussaint, G. (1995). *No quadrangulation is extremely odd*. Paper presented at the Sixth international symposium on algorithms and computation (formerly SIGAL International Symposium on Algorithms).
- Burges, C. J. C., Platt, J. C., & Jana, S. (2003). Distortion discrimination analysis for audio fingerprinting. *IEEE Transactions on Speech & Audio Processing*, 11(3), 165-174. <http://dx.doi.org/10.1109/TSA.2003.811538>
- 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).
- Chang, Y. C., Bergman, L., Castelli, V., Li, C. S., Lo, M. L., & Smith, J. R. (2000). The onion technique: indexing for linear optimization queries. *SIGMOD Rec.*, 29(2), 391-402. <http://dx.doi.org/10.1145/335191.335433>
- Chazelle, B. (1985). On the convex layers of a planar set. *IEEE Transactions on Information Theory*, 31(4), 509-517. <http://dx.doi.org/10.1109/TIT.1985.1057060>
- Dalal, K. (2004). Counting the Onion. *Random Struct Algorithms*, 24(2), 155-165. <http://dx.doi.org/10.1002/rsa.10114>
- Deliyannis, I. (2012). From Interactive to Experimental Multimedia. In I. Deliyannis (Ed.), *Interactive Multimedia* (pp. 3-12). Rijeka, Croatia: Intech. <http://dx.doi.org/10.5772/2394>
- Deliyannis, I., & Karydis, I. (2011). *Producing and Broadcasting Non-Linear Art-Based Content Through Open Source Interactive Internet-TV*. Paper presented at the ACM EuroITV.
- 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>
- Graham, R. L. (1972). An efficient algorithm for determining the convex hull of a finite planar set. *Information Process, Lett*, 1, 132-133. [http://dx.doi.org/10.1016/0020-0190\(72\)90045-2](http://dx.doi.org/10.1016/0020-0190(72)90045-2)
- Hess, W. (1983). *Pitch Determination of Speech Signals*. Heidelberg: Springer Verlag. <http://dx.doi.org/10.1007/978-3-642-81926-1>

- Kennedy, L., & Naaman, M. (2009). *Less talk, more rock: automated organization of community-contributed collections of concert videos*. Paper presented at the 18th International conference on World wide web.
- Kosch, H. (2004). *Distributed multimedia database technologies: supported by MPEG-7 and MPEG-21*: CRC Press.
- Marolt, M. (2008). A Mid-Level Representation for Melody-Based Retrieval in Audio Collections. *Multimedia, IEEE Transactions on*, 10(8), 1617-1625. <http://dx.doi.org/10.1109/TMM.2008.2007293>
- Milenkovic, V. (1999). Rotational polygon containment and minimum enclosure using only robust 2D constructions. *Computational Geometry*, 13(1), 3-19. [http://dx.doi.org/10.1016/S0925-7721\(99\)00006-1](http://dx.doi.org/10.1016/S0925-7721(99)00006-1)
- O'Rourke, J. (1998). *Computational Geometry in C*. New York: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511804120>
- Orio, N. (2010). Automatic identification of audio recordings based on statistical modeling. *Signal Processing*, 90(4), 1064-1076. <http://dx.doi.org/10.1016/j.sigpro.2009.10.025>
- Poulos, M., Belesiotis, V. S., & Papavaslopoulos, S. (2007). *Text identification via computational geometry*. Paper presented at the 6th WSEAS Conference on Artificial Intelligence, Knowledge Engineering and Databases.
- Poulos, M., Bokos, G., & Vaioulis, F. (2008). Towards the semantic extraction of digital signatures for librarian image-identification purposes. *JASIST*, 59(5), 708-718. <http://dx.doi.org/10.1002/asi.20753>
- Poulos, M., Magkos, E., & Chrissicopoulos, V. (2003). *Secure fingerprint verification based on image processing segmentation using computational geometry algorithms*. Paper presented at the International Conference on Signal Processing, Pattern Recognition and Applications (IASTED), Rhodes, Greece.
- Poulos, M., Papavaslopoulos, S., Bokos, G., Avlonitis, M., & Kanellopoulos, N. (2007). Specific Selection of FFT Amplitudes from Audio Sports and News Broadcasting for Classification Purposes. *Journal of Graph Algorithms and Applications*, 11(1), 277-307. <http://dx.doi.org/10.7155/jgaa.00146>
- Poulos, M., Papavaslopoulos, S., Chrissicopoulos, V., & Magkos, E. (2003). *Fingerprint verification based on image processing segmentation using an onion algorithm of computational geometry*. Paper presented at the Sixth Int Conf on Mathematics Methods in Scattering Theory and Biomedical Technology BIOTECH'6.
- Seifert, F. (2004). *Semantic music recognition audio identification beyond fingerprinting*. Paper presented at the Fourth International Conference in Web Delivering of Music, WEDELMUSIC. <http://dx.doi.org/10.1109/WDM.2004.1358108>
- Seo, J. S., Jin, M., Lee, S., Jang, D., Lee, S., & Yoo, C. D. (2005). *Audio fingerprinting based on normalized spectral subband centroids*. Paper presented at the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP'05. <http://dx.doi.org/10.1109/ICASSP.2005.1415684>
- 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>
- Stoica, P., & Moses, R. L. (1997). *Introduction to spectral analysis* (Vol. 51). Upper Saddle River, NJ: Prentice Hall.
- Sukittanon, S., & Atlas, L. E. (2002). *Modulation frequency features for audio fingerprinting*. Paper presented at the IEEE International Conference on Acoustics, Speech and Signal Processing ICASSP.
- 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>
- Turnbull, D., Barrington, L., Torres, D., & Lanckriet, G. (2008). Semantic Annotation and Retrieval of Music and Sound Effects. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(2), 467-476. <http://dx.doi.org/10.1109/TASL.2007.913750>
- Wang, A. (2003). *An industrial strength audio search algorithm*. Paper presented at the International Conference on Music Information Retrieval (ISMIR).