Multimed Tools Appl (2011) 51:881–895
DOI 10.1007/s11042-009-0420-7

Nonlinear dimensionality reduction for efficient and effective audio similarity searching

Dimitris Rafailidis · Alexandros Nanopoulos · Yannis Manolopoulos

Published online: 10 November 2009

Abstract In this paper, we address the issue of nonlinear dimensionality reduction to efficiently index spectral audio similarity measures. We propose the embedding of the spectral similarity space to a low-dimensional Euclidean space. This guarantees the triangular inequality and allows the adoption of several indexing schemes. We enlighten the advantages of the proposed indexable method against recently proposed spectral similarity measures that are also indexable. Moreover, our method compares favorably to linear dimensionality reduction methods, like multidimensional scaling (MDS). The proposed method significantly reduces the computation time during the construction process compared to any audio measure and, simultaneously, minimizes the searching cost for similar songs. To the best of our knowledge, the important issue of audio similarity measures' scalability is addressed for the first time.

Keywords Audio similarity searching · Content based retrieval · Music database · Nonlinear dimensionality reduction

1 Introduction

With the proliferation of available musical content, the development of musical content management methods is required to make available the millions of music titles to millions of users [2]. Music similarity measures have been proposed as a

D. Rafailidis (✉) · Y. Manolopoulos
Department of Informatics, Aristotle University, 54124 Thessaloniki, Greece
e-mail: draf@csd.auth.gr

Y. Manolopoulos
e-mail: manolopo@csd.auth.gr

A. Nanopoulos
Institute of Computer Science, University of Hildesheim, 31141 Hildesheim, Germany
e-mail: nanopoulos@ismll.de

means to search for relevant music. The development of music similarity measures has gained significant attention during the previous years [2, 4, 15, 16], and in particular, spectral music similarity measures, [2, 15].

A question of particular interest is whether existing music similarity measures can be successfully incorporated in real-world applications, such as search engines (e.g., music.yahoo.com) and recommender systems (e.g., last.fm, pandora.com). Two key factors that determine the successful application of similarity measures are: (i) their effectiveness in detecting perceptually relevant music, and (ii) their scalability in terms of their efficiency in searching large music collections.

In terms of the first above mentioned key factor (i.e. detecting perceptually relevant music), spectral-based similarity measures have helped in achieving good performance. It is worth noting that in the two events of 2007/08 of the Music Information Retrieval Evaluation eXchange,[1] extensions of spectral similarity measures have taken the first place for the task of Audio Music Similarity and Retrieval. Thus, it can be considered that this key factor has been satisfactorily investigated.

On the other hand, there is no analogous progress in terms of the second key factor (i.e. efficiency in combination with scalability), since currently proposed solutions do not guarantee good performance when searching large music collections. In principle, indexing strategies rely on the triangular inequality (i.e. given the audio distance measure $D(x, y)$ between a pair of songs x, y, for three songs x, y, z we may have $D(x, y) \not\leq D(x, z) + D(z, y)$) to prune the search space. However, most common spectral-based similarity measures (like G1C [16]) do not obey the triangular inequality and, thus, cannot be indexable. This fact recently motivated researchers to develop spectral-based similarity measures obeying the triangular inequality [12]. Nevertheless, although these new spectral-based similarity measures are indexable, they demonstrate rather poor performance as will be shown later. In addition, to the best of our knowledge, nonlinear dimensionality reduction has not been applied for audio indexing.

Motivated by the previous arguments, in the present paper our main goal is to efficiently index large music databases using spectral similarity measures. This is achieved in two steps: (a) by embedding audio similarity measures into a reduced space, and then (b) by indexing the reduced data with multidimensional index structures. This way, our methodŠs construction cost is dramatically reduced (10-15 times) in comparison to that of the previous methods, whereas the searching cost for similar songs is even more dramatically reduced (up to 25 times for large datasets). Thus, to the best of our knowledge, for first time in the literature, scalable efficient audio searching similarity has been achieved.

To meet this goal, we apply nonlinear dimensionality reduction (NLDR) techniques (such as L-Isomap and Local Linear Embedding) to embed the audio similarity space into a low-dimensional Euclidean space, where the triangular inequality holds, and several indexing strategies can be used. Our results demonstrate that the perceptual similarities in the original space have been preserved in a adequate way. We also show that the proposed approach outperforms the classic, linear methods, such as multidimensional scaling (MDS).

The rest of this paper is organized as follows. Section 2 describes the related work. In Section 3 we present the proposed approach, whereas Section 4 discusses

[1] http://www.music-ir.org/mirex

alternative multidimensional indexes and describes the mapping of query songs to be searched. Section 5 contains detailed experimental results and, finally, Section 6 concludes this paper.

2 Related work

Spectral music similarity measures have attracted research interest [2, 15]. They model timbre through the long-term distribution of local spectral features. Moreover, spectral measures have been combined with additional information, such as fluctuation patterns [16]. In particular, spectral music similarity measures consider audio signals as sequences of short overlapping frames. From each frame, spectral representation features are extracted (e.g., Mel Frequency Cepstrum Coefficients (MFCCs)). The overall distribution of features is summarized using a statistical model, such as clustering or Gaussian Mixture Model (GMM). The distance between two musical signals is computed by comparing their models with, e.g., Earth Movers Distance [15] or Monte Carlo sampling of the Kullback-Leibler Distance [2]. In the sequel, we focus on such spectral measures and their extensions.

The *Single Gaussian Combined* (G1C) audio distance measure [16] combines spectral similarity with *fluctuation patterns* that describe additional characteristics of the audio signal measures, like temporal information. Fluctuation patterns consist of three components: (i) the amplitude modulation of the loudness per frequency band, (ii) the sum of the values in the two lowest frequency bands with a modulation frequency higher than 1Hz, and (iii) the center of gravity of the fluctuation patterns on the modulation frequency axis.

For general spectral similarity measures, there is no guarantee that they obey the triangular inequality. As this is an important characteristic of metric spaces and indexing algorithms rely on it, Jensen et al. [12] proposed the use of the normalized L2 (Norm L2) distance. More specifically, they proposed prescaling all GMMs to have unit L2-norm and then considering the ordinary L2 distance between the scaled GMMs. Nevertheless, as will be shown later, the fact that the triangular inequality holds does not suffice to provide an efficient indexing scheme. Recently, Pohle and Schnitzer [17] proposed G1Cmod, which is a combination of G1C and normalized L2 distance. G1Cmod was ranked first in the MIREX 2007 contest (AudioSim evaluation task), as it slightly improves G1C. Due to the combination scheme, G1Cmod guarantees the triangular inequality.

Dimensionality reduction has been used for the visualization of music collections. An excellent source of relevant works for this topic can be found in [3]. However, to the best of our knowledge, nonlinear dimensionality reduction has not been applied to index music signals. Except for applications in music, dimensionality reduction methods have been applied in large time series databases [7, 13], where the high dimensionality of the data is a common problem. The offered solution is to perform dimensionality reduction on the data and then indexing the reduced data with a multidimensional index structure. A more general nonlinear dimensionality reduction technique for fast similarity search in large databases than the latter techniques can be found in [20].

When embedding in lower dimensional spaces in time-series databases, it is important to guarantee the preservation of proximities. However, in the case of music databases, music similarity measures (like G1C, G1Cmod etc) are subjective as they rely on the individual perception criteria that vary significantly among listeners. For these reasons, evaluation of similarity measures in music information retrieval is often performed with indirect objective measures, like precision in terms of finding results of the genre among the k nearest neighbors of a query (see for example the evaluation in works [1, 2, 12, 16, 17]. It is this characteristic of similarity measures in music information retrieval that, in contrast to the use of similarity measures in other fields (like in time-series), allows for not preserving exactly the same neighborhood into a transformed space.

3 Embedding the similarity space into a Euclidean space

Let $D(x, y)$ be a distance measure between a pair of songs x, y belonging to a collection of N songs. We assume that function D is non-negative and symmetric, and may not necessarily obey the triangular inequality. Our objective is to embed the objects into a d-dimensional Euclidean space $\mathbb{R}^d$, which well-approximates the dissimilarities in D.

The widely used method of Multidimensional Scaling (MDS) [5] (as well as the method of Principal Component Analysis) performs linear embedding. Its only assumption is the existence of a monotonic relationship between the original and the projected pair-wise distances. However, MDS may fail to well preserve the neighborhood of each song, i.e., songs that are close neighbors with respect to the original distance measure may not be neighbors in the embedded space.

For this problem several nonlinear algorithms have been devised, which are categorized either as *local* or *global*. Local linear embedding (LLE) [19] is a prominent local approach presented in Fig. 1a, whereas Isomap [9] is a global approach presented in Fig. 1b. (In both figures, parameter k represents the k most similar songs to a given song.) In Isomap, the lengths of edges in the neighborhood graph G provide a trustworthy guide to the local metric structure in the original space. The shortest-paths computation gives an estimate of the global metric structure, which can be provided into MDS to produce the required embedding.

LLE:

1. For each song x, find its k nearest neighbors (k-NNs) according to D.
2. Compute a weight vector $\overrightarrow{w_x}$ that best reconstructs x by a linear combination of its k-NNs.
3. Embed x to a point y by minimizing the reconstruction error of y using $\overrightarrow{w_x}$ and its corresponding k-NNs in $\mathbb{R}^d$.

(a)

Isomap:

1. For each song x, find its k nearest neighbors (k-NNs) according to D. Construct a neighborhood graph G. Connect x and each of its neighbors x' with an edge weighted by $D(x, x')$.
2. Compute the shortest paths in G for all pairs of songs. Store these distances in matrix Δ (i.e., $\Delta(i, j)$ is the geodesic distance between songs i, j).
3. Apply MDS on Δ to embed x to a point y in $\mathbb{R}^d$.

(b)

Fig. 1 Nonlinear dimensionality reduction algorithms (k is the number of nearest neighbors) (**a**, **b**)

Global algorithms, like Isomap, give more faithful representations of the data's global structure [9]. Isomaps complexity is $O(kN^2 \log N)$ for finding the shortest paths between all pairs (step 2) and $O(N^3)$ for MDS (step 3). Therefore, it is impractical to apply Isomap to large data sets. In contrast, LLE demonstrates an $O(dN^2)$ complexity, d being the dimensionality.

Landmark Isomap (L-Isomap) has been proposed to overcome this limitation [9]. In L-Isomap $n \ll N$ songs are randomly selected as landmarks, where $n > d + 1$. L-Isomap operates similarly to Isomap, however, only distances between all songs and the n landmark songs are preserved. This is achieved using a Landmark MDS procedure [9]. With L-Isomap, step 2 has complexity $O(knN \log N)$, whereas step 3 runs in $O(n^2 N)$. Thus, for $n \ll N$, the complexity of L-Isomap reduces to $O(N \log N)$.

4 Searching in indexing schemes

4.1 Data structures

In this section we describe ways to organize points in $\mathbb{R}^d$ (low-dimensional Euclidean space) using indexing schemes.

The M-tree is a balanced, dynamic tree data structure that partitions objects on the basis of their relative distances [8]. Leaves store all indexed objects, whereas internal nodes store *routing* objects. An example of M-tree is depicted in Fig. 2a ($C_1, \ldots, C_4$ are routing objects).

The R-tree family [11] consists of dynamic tree structures. The M-tree is general, as it requires only a distance function between the data. Differently, the R-tree can only index vectors in a Euclidean space, thus not any distance function can be used with it. Each entry within a non-leaf node stores a pointer to a child node and the minimum bounding rectangles of the child node. An example of R-tree is depicted in Fig. 2b.

Locality Sensitive Hashing (LSH) is a promising method of indexing high-dimensional data for k-NN queries [10]. The basic idea is to hash the input items to: (i) maximize the probability that similar items are mapped to the same buckets, and (ii) minimize the probability that dissimilar items are fallen to the same bucket. In general, LSH-based methods are approximate, which means that they do not necessarily return k results for a k NN query. The latter is measured by the *miss ratio* [10]. In our implementation we tune LSH to zero the miss ratio.

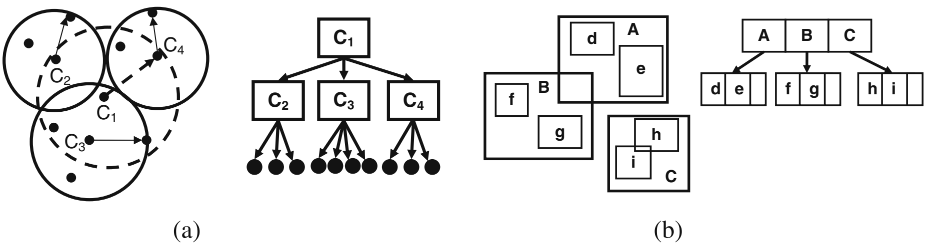


Fig. 2 Example of: **a** M-tree. **b** R-tree

4.2 Mapping the query song and similarity searching

Here, we first describe how to map a query song q to a d-dimensional vector in the $\mathbb{R}^d$ embedded space. Then, we describe how to search its nearest neighbors in this space.

For L-Isomap, we first apply classic MDS to the $n \times n$ original (spectral) distance between each pair of the n landmarks. Thus, we derive d eigenvectors $\mathbf{v_1}, \ldots, \mathbf{v_d}$ (each eigenvector being n-dimensional) and d eigenvalues $\lambda_1, \ldots, \lambda_d$. To map the query song q, let $\delta_{\mathbf{1}}, \ldots, \delta_{\mathbf{n}}$ be vectors of the original distance from each landmark to all other landmarks (thus, each $\delta_{\mathbf{i}}$ is n-dimensional as well). Also, let $\overline{\delta}$ be the mean of these vectors. For q, let $\widehat{\delta}$ contain the original distances between q and each landmark (thus, $\widehat{\delta}$ is n-dimensional too). The query song q is mapped to a d-dimensional vector $\mathbf{x} \in \mathbb{R}^d$, whose i-th coordinate is [9]:

$$\mathbf{x}_i = -\frac{1}{2} \frac{\mathbf{v_i}^T}{\sqrt{\lambda_i}} (\widehat{\delta} - \overline{\delta})$$

The above equation is equivalent to projecting onto the first d principal components of the landmarks [9]. For the case of LLE, an analogous mapping is described in [14].

Having mapped the query song to a point in $\mathbb{R}^d$, we use the index to search for its k-NN points. All indexes prune the search space and offer significant savings compared to sequential searching. More details on the searching algorithms can be found in [8, 10, 11].

5 Experimental results

5.1 Experimental settings

In this section, we will first compare nonlinear vs. linear dimensionality reduction methods, and we will show the advantage of the former ones. Next, we will examine how well the nonlinear methods preserve the geometry of the original similarity space in terms of precision for the task of audio similarity and retrieval (the baseline being the MDS method). In particular, to evaluate the effectiveness, for a given query song we will compute its k-NN songs and will measure the precision, i.e., how many of these k-NN songs belong to the same genre as the query song. Following recent research work [16, 17], we will apply the *artist filter*, i.e., we avoid having the same artist in both testing and training sets. Also, our measurements will be performed with leave-one-out cross validation. To address the genre imbalanced distribution problem (details in the next subsection), which is met in our data set, we will perform the Receiver Operating Characteristic curve (ROC) analysis.

Afterwards, we will compare the construction cost of G1Cmod algorithm to that of the proposed approach. In the sequel, we will compare the application of M-tree before and after dimensionality reduction, because Norm L2 and G1Cmod distance measure can be directly indexed with an M-tree (here, the baseline being the Norm L2 and G1Cmod method). We focus on M-tree, since it does not require a set of points as input; instead, it operates only with a distance matrix. Finally, we will show that the nonlinear methods admit dramatically more efficient implementations with several indexing schemes. To evaluate the efficiency, we will fist calculate the construction cost of G1Cmod as well as that of the proposed method by using as cost measure the number of distance computations required to find the k-NN songs [8]. To simplify the illustrations, we will give relative costs, which are normalized with respect to the cost of sequential searching (including the cost to embed the query song).[2]

We implemented all dimensionality reduction methods with the Dimensionality Reduction toolbox.[3] Regarding the audio similarity measures, we examined the Norm L2 and the G1Cmod method, which are indexable as they satisfy the triangular inequality [12, 17]. As default similarity measure for the dimensionality reduction methods, we also used the G1Cmod, whereas the Euclidean distance was the default distance measure in the embedded space.

We used the following collections: MIREX'04[4] and USPOP'02.[5] MIREX'04 (also called Magnatune), which has been used for the MIREX'04 genre classification contest, consists of 729 songs from 6 genres, performed by 128 artists. The files were downsampled to 22,050 Hz. USPOP'02 contains 8,764 songs from 10 genres, performed by 400 artists. The MFCCs are readily provided for this collection. In both sets, the framesize is 512 samples and the hopsize is 512 samples. 20 MFCCs are extracted from each frame and 30 sec from the center of each song are used. The computation of the audio similarity measure is done with the MA Toolbox[6] with required changes for the G1Cmod [17].

5.2 The imbalanced class distribution problem

Tables 1 and 2 present genres distribution in MIREX'04 and USPOP collection. It is necessary to devise a method to handle the imbalanced class distribution problem, since both collections consist of one dominated class-genre and many rare classes-genres. An efficient way to face this problem is a graphical method called the Receiver Operating Characteristic curve analysis (ROC) [18]. A ROC curve displays the tradeoff between the *true positive rate* (TPR) and the *false positive rate* (FPR). TPR is the fraction of positive examples predicted correctly, whereas FPR is the fraction of negative examples predicted as positive. In a ROC curve, TPR (FPR) is plotted along the y (x) axis. The ideal result is when the ROC curve is located as close as possible to the upper left corner of the diagram, whereas a random guess resides along the main diagonal. An alternative way to solve the class imbalanced problem is to estimate the area under the ROC curve (AUC) [6]. AUC equals to 1 (0.5) in case of a perfect result (random guessing). A method that is strictly better than another would have a larger AUC value.

[2] We have used the following index implementations: M-tree: www-db.deis.unibo.it/Mtree, R-tree: www.rtreeportal.org, LSH: www.cs.brown.edu/gregory/code/lsh.

[3] www.cs.unimaas.nl/l.vandermaaten/Laurens_van_der_Maaten

[4] http://www.music-ir.org/evaluation/m2k/release/README.htm#14

[5] http://labrosa.ee.columbia.edu/projects/musicsim/uspop2002.html

[6] http://www.ofai.at/~elias.pampalk/ma

Table 1 Genres distribution in MIREX '04

Genres	Classical	Electronic	Jazz & Blues	Metal & Punk	Rock & Pop	World
Percentage	43.5%	15.8%	3.6%	6.2%	14.1%	16.8%

Table 2 Genres distribution in USPOP

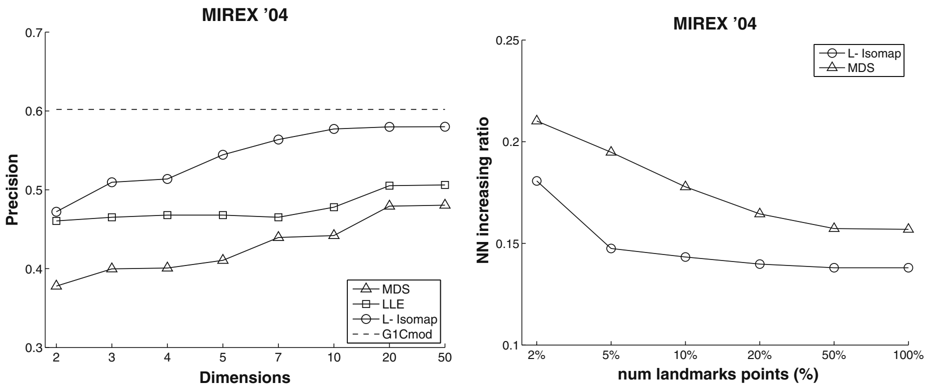
Genres	Country	Electronic	Jazz	Latin	New age	R&B	Rap	Reggae	Rock	Vocal
Percentage	2.9%	5.5%	0.5%	1.2%	0.7%	8.5%	5.2%	1.1%	74.1%	0.3%

5.3 Results

5.3.1 Precision maintenance

First, we compared dimensionality reduction methods: L-Isomap, LLE, and MDS. Since we are interested in methods that are both scalable and do not require the knowledge of query songs beforehand, we used a landmark MDS method [9]. Figure 3a shows a plot of precision against the dimensionality of the embedded space (the number of nearest neighbors is set to $k = 5$) for the MIREX'04 collection. The two nonlinear methods outperformed MDS, whereas L-Isomap presented the best performance. After 20 dimensions, the increase of dimensions presented only slight improvement in precision, but this improvement did not pay-off, as it required significant increase of the searching cost, since the dimensionality increase affected the index efficiency.

To evaluate the impact of the number of landmark songs, n, in L-Isomap, in Fig. 3b we present nearest neighbors increasing ratio against n (given as percentage of the collection size, N) for the MIREX'04 collection. The *nearest neighbors increasing ratio* shows how well we preserve the original distances in the embedded space. In particular, we validated if songs that are close neighbors with the original distance measure, are also close in the embedded space. For example, given a query P, the distance of his first neighbor P_1 in the original space is d_1. Then, we assume that his



(a) (b)

Fig. 3 Comparison of dimensionality reduction methods (**a**, **b**)

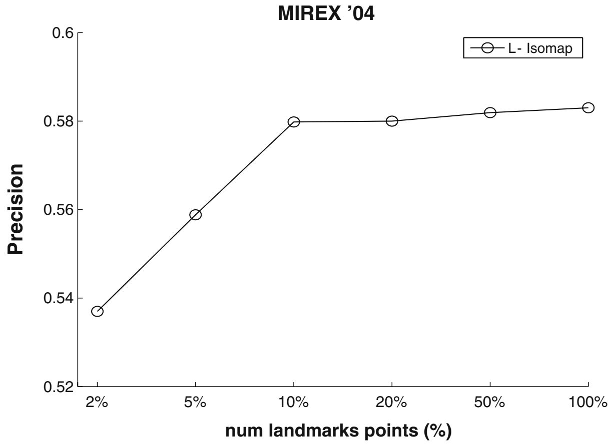


Fig. 4 L-Isomap precision against number of landmarks

first neighbor in the embedded space is P'_1, where the distance of P'_1 and P in the original space is d'_1. We compute the NN increasing ratio R as:

$$NN\ increasing\ ratio = \frac{|d_1 - d'_1|}{d_1}$$

Our goal was to maintain this ratio close to zero. A small n (e.g., 10%) sufficed as more landmarks did not improve the L-Isomap method. For $n \geq 10\%$, the increase of landmarks reduced slightly the increasing ratio, but this improvement did not pay-off, as it significantly increased the construction cost of the audio similarity measure. Additionally, MDS caused larger increasing ratio than L-Isomap and explained MDS' average behavior in terms of precision.

To ensure the appropriate landmarks points selection, we demonstrate in Fig. 4 how the increase from 2% to 100% can affect the L-Isomap method in terms of precision. As expected, for $n \geq 10\%$, the precision was slightly increased.

Next, we selected L-Isomap as the default dimensionality reduction method and compared its application against G1Cmod (without dimensionality reduction).

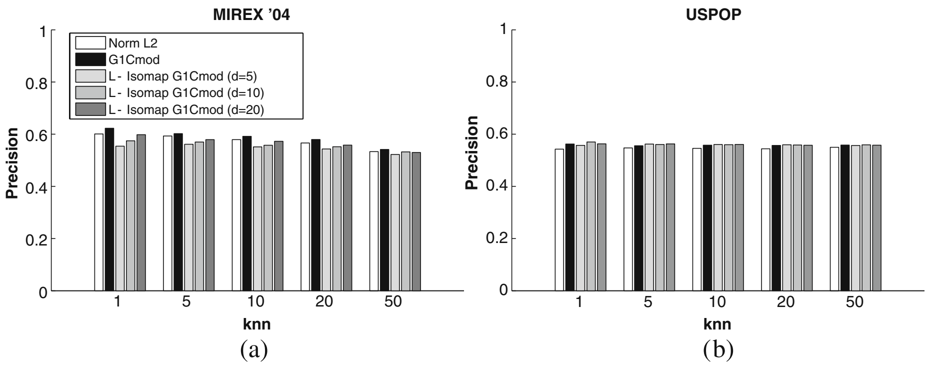


Fig. 5 Effectiveness results in terms of precision (**a**, **b**)

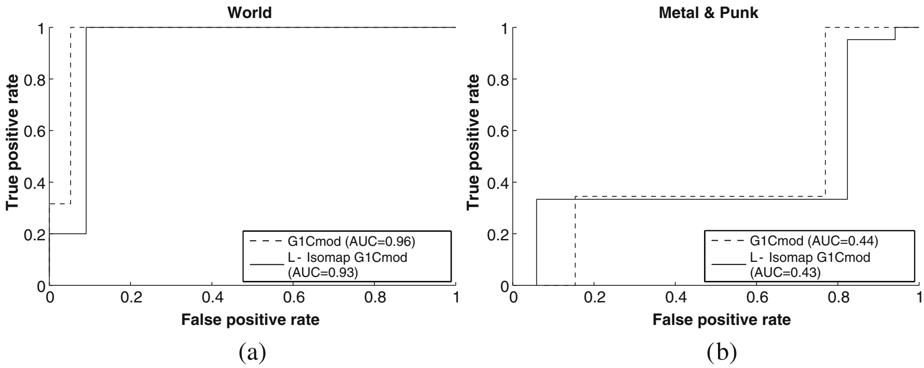


Fig. 6 ROC curves for MIREX '04 collection (**a**, **b**)

Figures 5a and b present the precision against the number k of nearest neighbors, for collections MIREX'04 and USPOP, respectively. For L-Isomap we examined 3 different dimensions: $d = 5, 10, 20$ and according to the previous experiments in Fig. 3 we considered $n = 10\%$ of landmark points. As observed, the precision of L-Isomap is comparable to the precision of G1Cmod and Norm L2, especially for higher dimension values. In particular the precisions have minimal differences in USPOP collection. To verify this, for both collections we applied statistical pairwise t-test; the calculated differences of means were insignificant at level 0.05. The dense number of songs in USPOP and the imbalanced genres distribution caused the invariable precision for different numbers of nearest neighbors.

According to the imbalanced class distribution problem, as observed in Tables 1 and 2, we performed a ROC analysis for MIREX'04 and USPOP. Figures 6 and 7 demonstrate the ROC curves with the respective AUC values for G1Cmod and L-Isomap G1Cmod. Instances of genres World and Metal & Punk in MIREX'04 represent 16.8% and 6.2% of the data set, whereas instances of genres New age and Country in USPOP represent 0.7% and 2.9%. In rare genres like these, L-Isomap maintained approximate values of AUC with AUC values of G1Cmod. For all genres

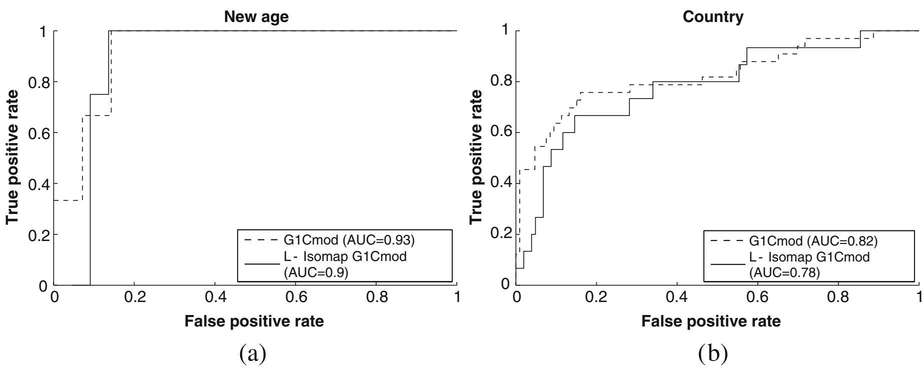


Fig. 7 ROC curves for USPOP collection (**a**, **b**)

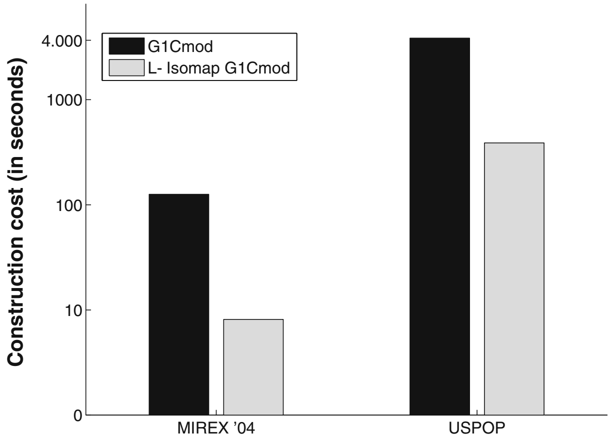


Fig. 8 Construction cost of audio similarity measure

of both collections we performed statistical pair-wise t-test on AUC values and the estimated differences of means were insignificant at level 0.05.

5.3.2 Cost optimization

Constructing the audio similarity measure of G1Cmod had high cost (quadratic complexity) because we had to compute the distances between all $\frac{N(N-1)}{2}$ songs, since G1Cmod is symmetric and thus we had to calculate only the upper triangular matrix. For L-Isomap G1Cmod we estimated the distances between $\frac{n(n-1)}{2}$ songs (n being a small fraction of N) and additionally we calculated the cost for the three steps of Isomap (see Fig. 1b), a cost with complexity $O(N \log N)$. In Fig. 8 we show how L-Isomap G1Cmod can reduce the respective cost of original G1Cmod, especially for the larger collection of USPOP (the y axis in Fig. 8 is in logarithmic scale). We note that the time costs for extracting the audio features are omitted, since they are identical for both methods. This dramatic performance improvement stems from the fact that practically the two approaches have different complexities.

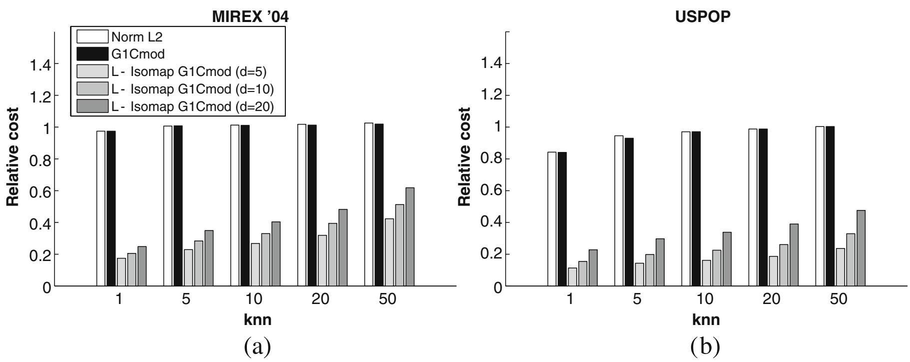


Fig. 9 Efficiency results in terms of relative cost (**a**, **b**)

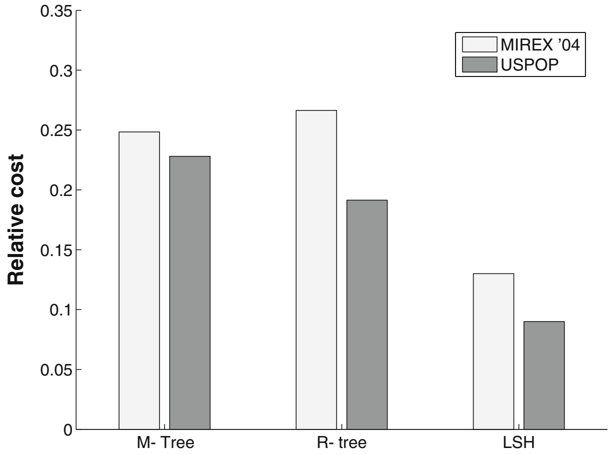


Fig. 10 Comparison of indexing schemes

To evaluate the efficiency of the searching process, we first used the M-tree to index the application of L-Isomap over G1Cmod. We also used the M-tree directly to Norm L2 and to G1Cmod, since they guarantee the triangular inequality. Figures 9a and b, for MIREX’04 and USPOP, respectively, present the relative cost (number of comparisons), normalized against the cost of sequential searching. Clearly, the M-tree created with L-Isomap outperforms Norm L2 and G1Cmod. It is worth noting that for large values of k, Norm L2 and G1Cmod costed slightly higher than the sequential algorithm. The reason is that in these cases they cannot filter the search space, plus they introduce additional cost for index searching.

Since L-Isomap produces multi-dimensional points in a Euclidean space, except the M-tree, we can examine the R-tree and LSH. Figure 10 presents (relative) costs for the case where we index the result of L-Isomap with the M-tree, R-tree, and LSH. All methods present substantial improvement compared to sequential searching in terms of speed-up factors. For instance, M-tree presents up to 6 and 9 times improvement, R-tree up to 8 and 13, and LSH up to 12 and 23, for MIREX’04 and USPOP respectively. LSH presents substantial improvement, especially for larger collections (USPOP’02). The reason is that LSH has been designed to work well in high dimensional spaces.

6 Conclusions

Previous efforts on audio similarity searching failed to provide efficient solutions in large music collections. In brief, our proposed scheme (a) adopts nonlinear dimensionality reduction techniques, and (b) uses multidimensional indexing schemes. Through a detailed generalized experimentation, we have examined several factors to tune all the methods under investigation. The results prove that our proposed scheme outperforms by far the previous methods achieving a robust scalable behaviour in the case of large music databases. Summarizing:

– the proposed approach is effective, because it perceptually preserves the geometry of the original similarity space.

– the proposed approach is efficient with respect to the construction and the searching cost, as it dramatically reduces the cost observed by the previous methods.

As future work, we will evaluate the proposed method in evolving music collections, where continuous query processing is required (data steams).

References

1. Aucouturier J-J (2006) Ten experiments on the modeling of polyphonic timbre. PhD thesis, University of Paris 6, France
2. Aucouturier J-J, Pachet F (2004) Improving timbre similarity: how high is the sky? Journal on Negative Results in Speech and Audio Sciences 1:1–11
3. Baumann S (2005) Music similarity and visualization for music IR. In: Tutorial in 6th international symposium on music information retrieval (ISMIR), London
4. Berenzweig A, Logan B, Ellis D, Whitman B (2003) A large-scale evaluation of acoustic and subjective music similarity measures. Comput Music J 28:63–76
5. Borg I, Groenen P (2005) Modern multidimensional scaling: theory and applications. Springer-Verlag, New York
6. Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recogn 30:1145–1159
7. Chakrabarti K, Keogh E, Mehrotra S, Pazzani M (2002) Locally adaptive dimensionality reduction for indexing large time series databases. ACM Trans Database Syst 27:188–228
8. Ciaccia P, Patella M, Zezula P (1997) M-tree: an efficient access method for similarity search in metric spaces. In: Proceedings of 23rd international conference on very large data bases (VLDB). Athens, Greece, pp 426–435
9. de Silva V, Tenenbaum JB (2002) Global versus local methods in nonlinear dimensionality reduction. In: Proceedings of neural information processing systems (NIPS), Vancouver, pp 705–712
10. Gionis A, Indyk P, Motwani R (1999) Similarity search in high dimensions via hashing. In: Proceedings of 25th international conference on very large data bases (VLDB), Edinburgh, pp 518–529
11. Guttman A (1984) R-Trees: a dynamic index structure for spatial searching. In: Proceedings of ACM international conference on management of data (SIGMOD), Boston, pp 47–57
12. Jensen JH, Ellis D, Christensen MG, Jensen SH (2007) Evaluation distance measures between Gaussian mixture models of MFCCs. In: Proceedings of 8th international symposium on music information retrieval (ISMIR), Vienna, pp 107–108
13. Keogh E, Chakrabarti K, Pazzani M, Mehrotra S (2001) Dimensionality reduction for fast similarity search in large time series databases. Knowl Inf Syst 3:263–286
14. Kouropteva O, Okun O, Pietikäinen M (2005) Incremental locally linear embedding. Pattern Recogn 38:1764–1767
15. Logan B, Salomon A (2001) A music similarity function based on signal analysis. In: Proceedings of IEEE international conference on multimedia and expo (ICME), Tokyo, pp 745–748
16. Pampalk E, Flexer A, Widmer G (2005) Improvements of audio-based music similarity and genre classification. In: Proceedings of 6th international symposium on music information retrieval (ISMIR), London, pp 628–633
17. Pohle T, Schnitzer D (2007) Striving for an improved audio similarity measure. MIREX 2007-audio music similarity and retrieval (http://www.music-ir.org/mirex/2007/)
18. Provost FJ, Fawcett T (1997) Analysis and visualization of classifier performance: comparison under imprecise class and cost distributions. In: Proceedings of international conference on knowledge discovery and data mining (KDD), Newport Beach, pp 43–48
19. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290:2323–2326
20. Vu K, Hua LA, Cheng H, Lang S (2006) A non-linear dimensionality-reduction technique for fast similarity search in large databases. In: ACM international conference on Management of Data (SIGMOD), Chicago, pp 527–538

Dimitris Rafailidis received a B.Sc. (2005), and a M.Sc. (2007) in Information Systems in Computer Science Department of the Aristotle University of Thessaloniki. Currently he is a PhD student in the same institute. His research interests include databases, data mining and Web information systems.

Alexandros Nanopoulos is a Professor of Machine Learning in Hildesheim University, Germany. He obtained his PhD from the Aristotle University of Thessaloniki, Greece. The title of his dissertation was "Techniques for Non Relational Data Mining". From 2004 to 2008 he has been a Lecturer in Aristotle University teaching data mining. Alexandros Nanopoulos is co-author of more than 70 articles in international journals and conferences. He has also co-authored the monograph "Advanced Signature Techniques for Multimedia and Web Applications and "R-trees: Theory and Applications". His research interests include data mining, machine learning, information retrieval and databases.

Yannis Manolopoulos received the BEng degree (1981) in electrical engineering and the PhD degree (1986) in computer engineering, both from the Aristotle University of Thessaloniki. Currently, he is a professor in the Department of Informatics at the same university. He has been with the Department of Computer Science at the University of Toronto, the Department of Computer Science at the University of Maryland at College Park, and the University of Cyprus. He has published more than 200 papers in journals and conference proceedings. He is coauthor of four monographs: "Advanced Database Indexing", "Advanced Signature Indexing for Multimedia and Web Applications" (both by Kluwer), "R-Trees: Theory and Applications", "Nearest Neighbor Search: A Database Perspective" (both by Springer). He has coorganized several conferences (among others ADBIS'02, SSTD'03, SSDBM'04, ICEIS'06, ADBIS'06, and EANN'07). His research interests include databases, data mining, Web information systems, sensor networks, and informetrics.