International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:12, 2008

Application of a Novel Audio Compression scheme in Automatic Music Recommendation, Digital Rights Management and Audio Fingerprinting

Anindya Roy, and Goutam Saha

***Abstract*—**Rapid progress in audio compression technology has contributed to the explosive growth of music available in digital form today. In a reversal of ideas, this work makes use of a recently proposed efficient audio compression scheme to develop three important applications in the context of Music Information Retrieval (MIR) for the effective manipulation of large music databases, namely automatic music recommendation (AMR), digital rights management (DRM) and audio finger-printing for song identification. The performance of these three applications has been evaluated with respect to a database of songs collected from a diverse set of genres.

***Keywords*—**Audio compression, Music Information Retrieval, Digital Rights Management, Audio Fingerprinting.

I. Introduction

RECENT advances in audio coding techniques has led to the phenomenal growth in the number of available digital audio files. Today, the average listener is able to conveniently access a large database of music, spanning a wide spectrum of genres and artists. However, it is difficult to effectively organize, search and manipulate such a large database, leading to the development of a new field of research, viz. Music Information Retrieval [1]. Further, the ease of duplicating music in digital form has lead to the problem of copyright protection and digital rights management. In this paper, three separate applications related to the efficient handling of such large music databases have been proposed, utilizing the **Q**uantization-optimized multilevel **K**arhunen-Loève Transform (QK) algorithm, an audio compression algorithm developed by the authors [2] based on a generalization of the Karhunen-Loève Transform across multiple levels.

Manuscript received May 1, 2007.

Anindya Roy is with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, Kharagpur-721302, India (phone: +91 98301 39279; e-mail: anindya.ici@ gmail.com).

Goutam Saha is with the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, Kharagpur-721302, India (phone: +91-3222-283556; e-mail: gsaha@ ece.iitkgp.ac.in).

The choice of a listener is often decided by triple factors, namely, the *bandwidth* available, the *quality* available at that bandwidth and the *genre* preferred by the listener. The recently proposed QK audio compression scheme [2] offers a way to combine the trade-offs incurred by these three separate factors. The performance of this scheme shows that, different genres of music exhibit strikingly different audio quality at the same bit rate. Here PSNR has been used as an objective measure of audio quality. However, this is not a constraint and a psycho-acoustic measure is expected to give a similar conclusion. Based on this observation, this work proposes an Automatic Music Recommendation (AMR) system [3]-[5] which presents the listener with a *well-defined* choice of genres, and a choice of songs within a particular genre if required, at the maximum available bit rate and minimum user permissible audio quality.

Further, the structure of the compression scheme makes it very suitable for effective protection of copyright in case of the compressed audio. Developing on this aspect, an audio database system has been proposed. This system views a music archive in a cryptographic context, leading to fast processing and efficient digital rights management [6]-[8] of encoded files.

Finally, with the plethora of available songs, it is of major importance to correctly identify a given instance of a song as well as possible duplicates. This is the field of audio fingerprinting [9]-[11]. This paper shows how the proposed QK audio compression scheme can be used to extract a reliable fingerprint of the audio. The fingerprinting scheme has been evaluated on a database of sixteen audio files derived from a diverse set of genres. In the results section, it has been shown that our scheme performs significantly well at identifying the correct audio file.

The rest of the paper is arranged as follows. Section II gives a theroretical background to the present work and a brief overview of the recently proposed QK audio compression algorithm, Sections III, IV and V details the three applications of this algorithm for handling large music

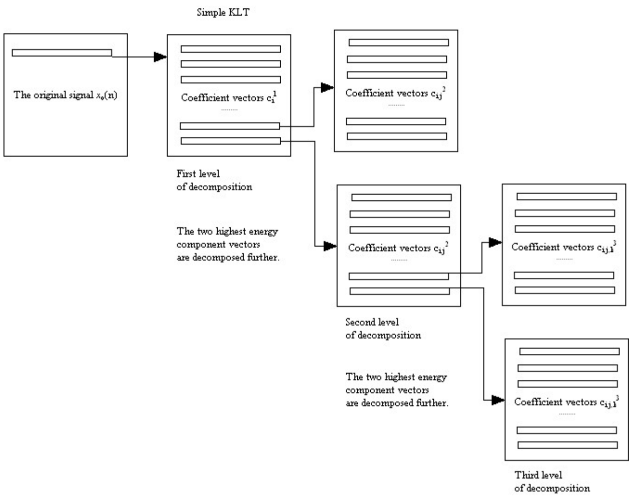

Fig. 1 The Multilevel Karhunen Loève Transform. *At each level, the first few highest energy components are further decomposed, leading to shorter and shorter sub-signals with more and more energy.* This leads to better utilization of the Energy Compaction property of the KL Transform

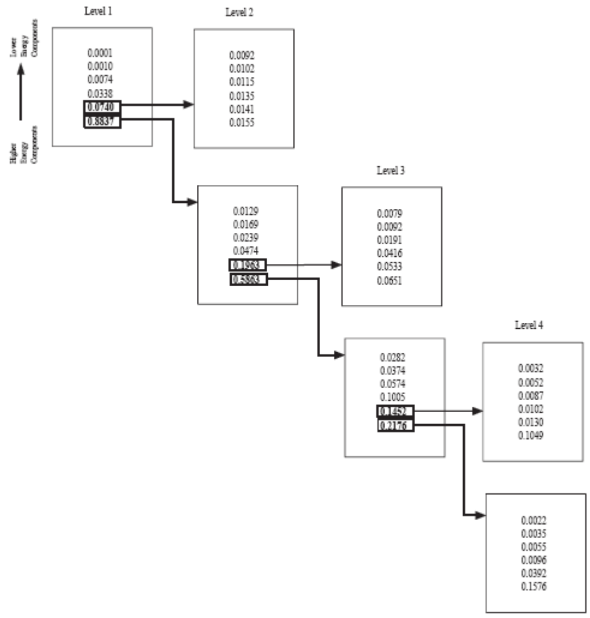

Fig. 2 Energy Compaction performance of the mKLT with respect to one music file, "Thank You - Dido". *At each level, more and more energy is concentrated in the first few highest energy subsignals, leading to better utilization of the energy compaction property of the KL Transform*

databases and in MIR, which is the chief contribution of this work. Finally, Section VI draws the principal conclusions of this work and directions for future research.

II. Theoretical Background

Before introducing the three proposed applications in MIR, a brief overview of the QK audio compression algorithm [2] is presented which gives the necessary theoretical background to proposed framework.

A. Notational Preliminaries

Throughout the paper, variable names in italics, for example, x_0, x_1, x_i, etc. indicate one-dimensional discrete time signals. Names in bold indicate vectors or the corresponding time signals broken up into non-overlapping blocks of a pre-specified length M. For example, $\mathbf{x}_{0,k}$, $\mathbf{x}_{1,k}$ and $\mathbf{x}_{i,k}$ denote the k-th block of length M extracted from the corresponding signals x_0, x_1 and x_i respectively. Finally, the subscript q beside a name indicates quantization in some form. For example, $\mathbf{c}_{q,i}$ denotes the quantized version of $\mathbf{c}_i$, $x_{q,i}(n)$ is the reconstructed signal after *lossy* compression of $x_i(n)$, while $\mathbf{e}_{q,i}$ is the *quantization optimized* version of the eigenvector $\mathbf{e}_i$.

B. The Karhunen-Loève Transform

The KL transform decomposes a signal into its principal components [12]-[15], [22]-[24]. It achieves maximum decorrelation of the signal components and minimum reconstruction error when fewer components are selected for reconstruction from the total. Further, it maximizes the Coding Gain CG [25], defined by,

$$CG = \log_{10}[(1/M)\sum_{i=0}^{M-1}\sigma_i^2 / (\prod_{i=0}^{M-1}\sigma_i^2)^{1/M}] \qquad (1)$$

upon compression by optimal bit allocation [26], where σ^2_i represents the variance of the i-th signal component and M is the total number of components.

Let $\{x_0(\mathrm{n})\}_{\mathrm{n}=1}^{\mathrm{N}}$ be the signal to be compressed. The signal $x_0(n)$ is broken up into non-overlapping blocks $\{\mathbf{x}_{0,k}\}_{k=1}^{N/M}$ of length M as,

$$\mathbf{x}_{0,k} = [\, x_0(Mk-M+1)\; x_0(Mk-M+2)..x_0(Mk-1)\; x_0(Mk)]^T \qquad (2)$$

where $1 \le (N/M) \le k$ and it is assumed that N is an exact multiple of M for mathematical simplicity.

As in practical applications, it is assumed that $\{\mathbf{x}_{0,k}\}$ is a zero mean random process, which facilitates mathematical simplicity without compromising on generality. It is required to find a unit magnitude M-vector $\mathbf{e}_1$ such that the magnitude of the projection vector $\mathbf{c}_1$ will be maximized in statistical sense, where $\mathbf{c}_1 = [c_1(1)\; c_1(2) \ldots c_1(\,N/M\,)]^T$ and

$$c_1(k) = \mathbf{e}_1^{\mathrm{T}}\mathbf{x}_{0,k} \text{ for } k = 1, 2, \ldots, (N/M) \qquad (3)$$

are the projections of $\{\mathbf{x}_{0,k}\}$ along the direction $\mathbf{e}_1$. Conversely, if the $\{\mathbf{x}_{0,k}\}$ are reconstructed using only the components $\{c_1(k)\}_{k=1}^{N/M}$, then this should give rise to minimum reconstruction error.

Mathematically, it is required to find $\mathbf{e}_{opt}$ such that the Error Criterion,

$$J_{KLT}(\mathbf{e}_1) = \sum_{k=1}^{N/M} \|\, \mathbf{x}_{0,k} - c_1(\mathrm{k})\mathbf{e}_1 \,\|^2 \qquad (4)$$

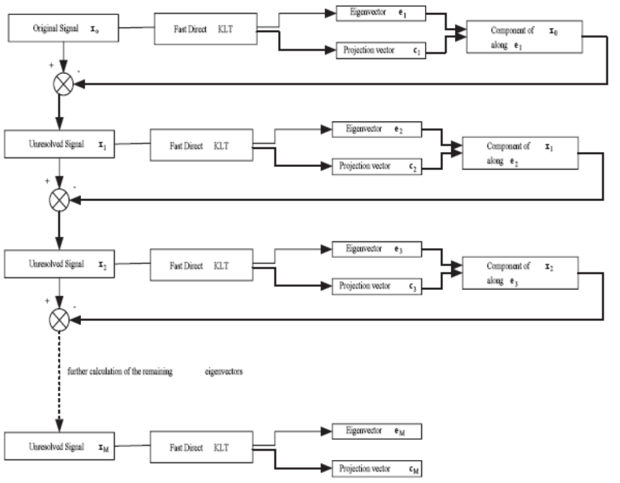

Fig. 3 The step-by-step implementation of the KL Transform with feedback between the steps

is minimized where $c_1(k) = \mathbf{e}_1^T\mathbf{x}_{0,k}$ (3). It can be easily shown that $\mathbf{e}_1$ satisfies the equation,

$$\mathbf{S}\mathbf{e}_1 = \lambda\mathbf{e}_1 \qquad (5)$$

i.e., $\mathbf{e}_1$ is an eigenvector of the Scatter Matrix

$\mathbf{S} = \sum_{k=1}^{N/M} \mathbf{x}_{0,k}\mathbf{x}_{0,k}^T$ of $\{\mathbf{x}_{0,k}\}$. Further, $\mathbf{e}_1$ should correspond to the largest eigenvalue $\lambda = \lambda_1$ of $\mathbf{S}$.

This choice of $\mathbf{e}_1$ gives rise to least reconstruction error for a particular order. Further, it can be shown that if $\{\mathbf{x}_{0,k}\}$ is projected along the M different eigenvectors $\{\mathbf{e}_i\}_{i=1}^M$ of $\mathbf{S}$, leading to M different projection vectors $\{\mathbf{c}_i\}_{i=1}^M$, maximum decorrelation and maximum Coding Gain can be acheieved, after optimal bit allocation (1).

In brief, the Karhunen-Loève Transform is expressed as a decomposition of the original signal $\{x_0(n)\}_{n=1}^N$ into M eigenvectors $\{\mathbf{e}_i\}_{i=1}^M$ and M subsignals $\{\mathbf{c}_i\}_{i=1}^M$ as follows,

$$KLT_{simple}[x_0(n)_{n=1}^N] = [\{e_i\}_{i=1}^M, \{c_i\}_{i=1}^M] \qquad (6)$$

C. Optimal Bit Allocation

After a signal has been decomposed into a certain number of subbands, say M, and an overall bit rate of R bits/sample is specified, an optimal bit allocation [26] for the different subbands can be calculated based on their respective subband variances, $\{\sigma_k^2\}_{k=1}^M$, as defined by the equation,

$$R_k = R + \frac{1}{2}\log_2 \frac{\sigma_k^2}{\prod_{j=1}^{M}(\sigma_j^2)^{1/M}} \qquad (7)$$

for $k = 1, 2, 3, \ldots, M$ where R_k is the number of bits allocated to the k-th subband. It is to be noted that in case of a decomposition based on the KLT, $\sigma_k^2 = \lambda_k$ where λ_k is the eigenvalue corresponding to the k-th component or subband, as obtained in Section II-B.

D. The Multilevel Karhunen-Loève Transform

In this phase, the simple KL Transform has been extended to multiple levels. By decomposing the highest energy components further into their principal components, it was hypothesized that we would be able to better utilize the property of energy compaction of the KL Transform. Consequently, it was found that this extension led to markedly better compression performance, justifying our hypothesis.

Let $\{x_0(\mathrm{n})\}_{\mathrm{n}=1}^{\mathrm{N}}$ be the signal to be compressed. The signal $x_0(n)$ is broken up into non-overlapping blocks $\{\mathbf{x}_{0,k}\}_{k=1}^{N/M}$ of length M as,

$$\mathbf{x}_{0,k} = [\, x_0(Mk-M+1)\; x_0(Mk-M+2)..x_0(Mk-1)\; x_0(Mk)]^T \qquad (8)$$

where $1 \le (N/M) \le k$ and it is assumed that N is an exact multiple of M for mathematical simplicity.

Using the simple KL Transform, the principal directions, $\{\mathbf{e}_i\}_{i=1}^M$ of $\{\mathbf{x}_{0,k}\}$ are calculated and $\{\mathbf{x}_{0,k}\}$ is projected along these M different eigenvectors $\{\mathbf{e}_i\}_{i=1}^M$, leading to M different projection vectors $\{\mathbf{c}_i\}_{i=1}^M$. However, unlike the simple KL Transform, the algorithm does not stop at this stage. In the crucial next step which forms the basis of our multilevel generalization of the KL Transform, a predefined number *nFertile* of projection vectors is set to be *further* decomposed. It is assigned $\{\mathbf{c}^1_i\}_{i=1}^M = \{\mathbf{c}_i\}_{i=1}^M$, and the first *nFertile* highest energy components, $\{\mathbf{c}^1_i\}_{i=1}^{nFertile}$ is extracted from this set. The KL Transform is applied in turn on these projection vectors, taking *themselves* now as new signals, as with $\{x_0(n)\}_{n=1}^N$ in the first step. Thus, the next level of projection vectors, $\{\mathbf{c}^2_{i,j}\}_{j=1}^M$ and eigenvectors $\{\mathbf{e}^2_{i,j}\}_{j=1}^M$ for $1 \le i \le nFertile$ is found out as follows,

$$KLT_{simple}[c_i^1(n)_{n=1}^{N/M}] = [\{e^2{}_{i,j}\}_{j=1}^M, \{c^2{}_{i,j}\}_{j=1}^M] \qquad (9)$$

for $1 \le i \le n_{Fertile}$. It should be noted that the length of the projection vectors are reducing by a factor of M at each level.

In subsequent steps, further levels of projection vectors, $\{\mathbf{c}^Z_{i,j,k....etc.}\}_{j=1}^M$ and eigenvectors, $\{\mathbf{e}^Z_{i,j,k....etc.}\}_{j=1}^M$ are calculated where Z is the predefined number of levels that has been set in advance. $Z = 1$ corresponds to the simple KL Transform. An illustrative flowchart of the mKLT has been provided in fig.1. After this decomposition process, a tree of sub-signals is obtained through which we can better utilize the energy compaction property of the KL Tranform, as shown in fig.2. Bits are allocated optimally to these sub-signals according to (7), to utilize this energy compaction property, leading to better compression performance.

E. Noise Removal Filter

This forms one important step for removal of Quantization Noise in the Synthesis stage of our technique and is based on the concept of the Wiener filter [27], [28]. In the given context,it is known in advance both the desired signal *x(n)* and the noisy signal $x_\eta(n)$, and hence also the noise signal $\eta(n) = x_\eta(n) - x(n)$. It is required to design a filter $\mathbf{w}$ of length L which can maximally reduce $\eta(n)$, allowing us to obtain the best estimate of $x(n)$ from $x_\eta(n)$ in a statistical sense. The length L shall be obtained from certain performance constraints explained in Section III.

International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:2, No:12, 2008

Let $\mathbf{x}_{\eta,k} = [x_\eta(k-B_1)\ x_\eta(k-B_1+1)\ \ldots\ x_\eta(k+B_2)]^T$ such that $B_1+B_2+1 = L$. Although a choice of $B_2 > 0$ makes the filter non-causal, this is not a major problem since a finite number of time advanced data will be used and output will be delayed by that amount.

It is required to find **w** such that the Error Criterion,

$$J_\eta(\mathbf{w}) = \sum_k \| \mathbf{w}^T\mathbf{x}_{\eta,k} - x(k) \|^2 \tag{10}$$

is minimized. With little mathematical exercise, it can be shown that such an optimal choice for **w** is given by,

$$\mathbf{w} = [\sum_k [\mathbf{x}_{\eta,k}\ \mathbf{x}_{\eta,k}^T]\ ]^{-1}[\sum_k x(k)\ \mathbf{x}_{\eta,k}] \tag{11}$$

F. The Quantization Optimized Karhunen-Loève Transform

The concepts developed in the preceding sections have been used to develop the final Quantization Optimized Karhunen-Loève Transform. In the usual Karhunen-Loève Transform [12]-[15] based compression schemes, the analysis or decomposition stage is not influenced by the subsequent quantization of the subsignals. In the algorithm recently proposed by the authors [2], the underlying philosophy is to start reducing quantization noise as soon as it is created, before it accumulates. Hence, as soon as the first component (i.e. the highest energy component, $\mathbf{c}_1$ in Section II-B) of the signal $x_0(n)$ has been calculated (Fig.3), bits are allocated optimally to it (Section II-C). Thus the quantized stream $\mathbf{c}_{q,1}$ is obtained from $\mathbf{c}_1$ and a best estimate, $x_{q,0}(n)$ of the original signal $x_0(n)$ (in the least square sense) is obtained, based on this (quantized) highest energy component $\mathbf{c}_{q,1}$ only. The difference between this estimate and the original signal, denoted as $x_1(n)$, which contains the contribution from quantization noise apart from the unresolved components of $x_0(n)$, is calculated. Next, KLT-based decomposition is carried out as before on this difference signal $x_1(n)$ just as with the original signal $x_0(n)$ and the next highest energy component $\mathbf{c}_{q,2}$ and its quantized version $\mathbf{c}_{q,2}$ is calculated. This process is repeated, yielding $\mathbf{c}_3$, $\mathbf{c}_{q,3}$, $\mathbf{c}_4$, $\mathbf{c}_{q,4}$ and so on, until certain performance parameters are exceeded. This step-by-step procedure for calculation of KL Transform has been illustrated in Fig. 3, while in Fig. 4, the incorporation of the quantization step in between two stages of the KL Transform has been shown.

Further, a nested tree-like multilevel decomposition structure is effected by decomposing the best few components $\{\mathbf{c}_i\}$ in the same way as the original signal at the time of their quantization, as shown in Fig. 1. This helps to maximally utilize the energy compaction property of the KL Transform.

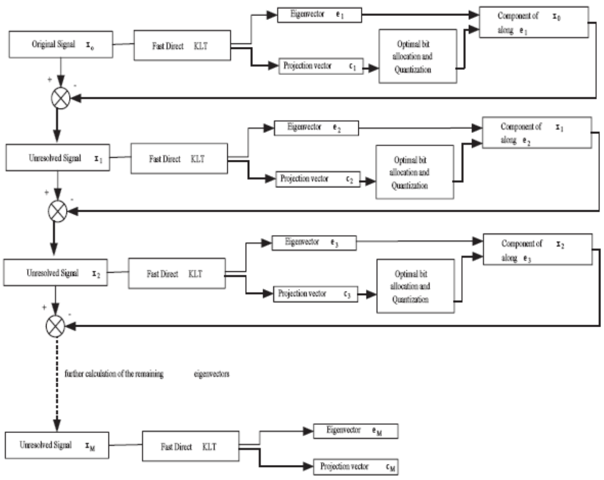

Fig. 4 The step-by-step KL Transform with quantization incorporated between intermediate steps

In our experiments, the first *two* components were decomposed, which gave satisfactory results, at the same time keeping the eigenvector overhead within reasonable limits.

Let $\{x_0(n)\}_{n=1}^{N}$ be the signal to be compressed. As in Section II-B, the signal $x_0(n)$ is broken up into non-overlapping blocks $\{\mathbf{x}_{0,k}\}_{k=1}^{N/M}$ of length M as,

$$\mathbf{x}_{0,k} = [x_0(Mk-M+1)\ x_0(Mk-M+2)\ \ldots\ x_0(Mk-1)\ x_0(Mk)]^T \tag{12}$$

where $1 \le k \le N/M$. Let R be the allowable bit rate set by the user.

The algorithm can be described in three steps as follows :-

Step 1

A] Applying a simple M-band KLT over $\{\mathbf{x}_{0,k}\}_{k=1}^{N/M}$, the principal eigenvector as $\mathbf{e}_1$ (ref. Section II-B) is obtained. For our application on audio files, an optimal value of $N = 16000$ and $M = 8$ was chosen from experimental studies in order to get best compression performance. At the same time, such a low value of M helps to limit the computational complexity of our algorithm.

B] Bits R_i are optimally allocated to each band i, from R, according to the formula in Section II-B using $\sigma_i^2 = \lambda_i$ where λ_i is the i-th highest eigenvalue of the KLT in Step 1A.

C] The $\{\mathbf{x}_{0,k}\}$ are resolved *only* along the best eigen-direction, $\mathbf{e}_1$, as obtained in Step 1A, finding a projection vector $\mathbf{c}_1$ as in (3), defined as, $\mathbf{c}_1 = [c_1(1)\ c_1(2)\ \ldots\ c_1(\ N/M\ )]^T$ and $c_1(k) = \mathbf{e}_1^T\mathbf{x}_{0,k}$. The remaining directions are ignored. They will be taken care of later.

D] This best coefficient vector $\mathbf{c}_1$ is quantized using Max-Lloyd Quantizer getting $\mathbf{c}_{q,1}$, according to the bits allocated to it in Step 1B. The resulting bit stream is entropy coded. To effect the nested multilevel structure, $\mathbf{c}_1$ can be further compressed following this same algorithm, leading to better utilization of the energy compaction property of the KL Transform.

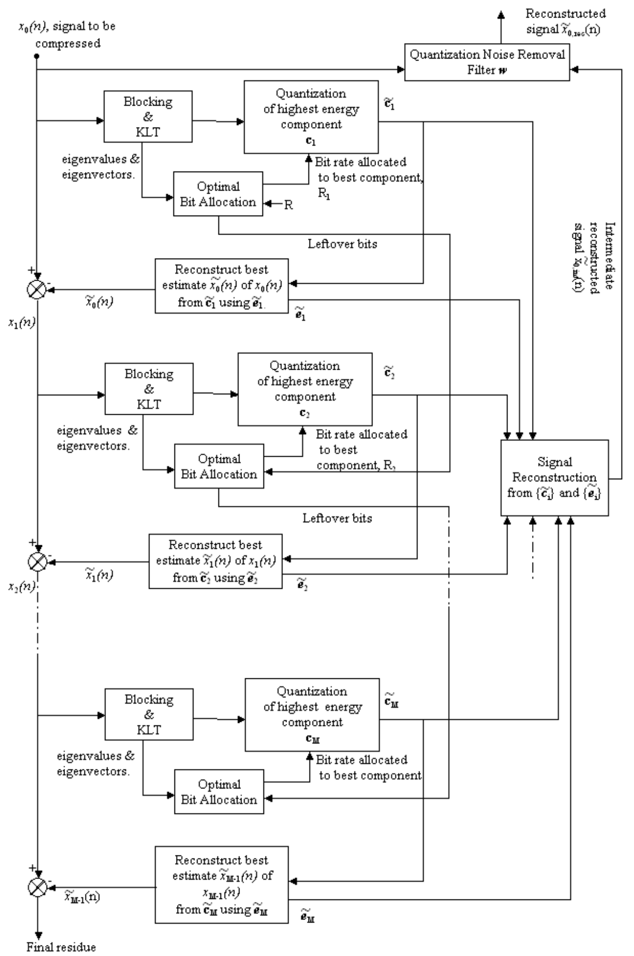

Fig. 5 Flowchart of the proposed Compression Scheme. Some of the higher energy quantization blocks may include a complete compression block within themselves, affecting the nested multilevel structure

E] The best vector $\mathbf{e}_{q,1}$ which can optimally reconstruct $x_0(n)$ from $\mathbf{c}_{q,1}$ is found. In general, $\mathbf{e}_1$ and $\mathbf{e}_{q,1}$ are not the same since $\mathbf{e}_{q,1}$ is calculated from the quantized version of $\mathbf{c}_1$, namely $\mathbf{c}_{q,1}$, in a feedback process while $\mathbf{e}_1$ was calculated directly from $x_0(n)$. This $\mathbf{e}_{q,1}$ is calculated using a simple distortion minimization criteria similar to Section II-C. The idea is that $\mathbf{e}_{q,1}$ will be better able to accommodate the quantization noise introduced in Step 1D than just $\mathbf{e}_1$.

F] An approximation $x_{q,0}(n)$ of $x_0(n)$ is reconstructed from $\mathbf{c}_{q,1}$ and $\mathbf{e}_{q,1}$.

Step 2

A] The residue signal $x_1(n)$ is obtained as, $x_1(n) = x_0(n) - x_{q,0}(n)$. Now this residue is treated as *approximation noise*, which should be taken care of in the succeeding steps. This step forms the feedback loop structure unique to this algorithm (Fig.3). Thus, the quantization noise introduced in previous stages influences subsequent stages so that all the stages

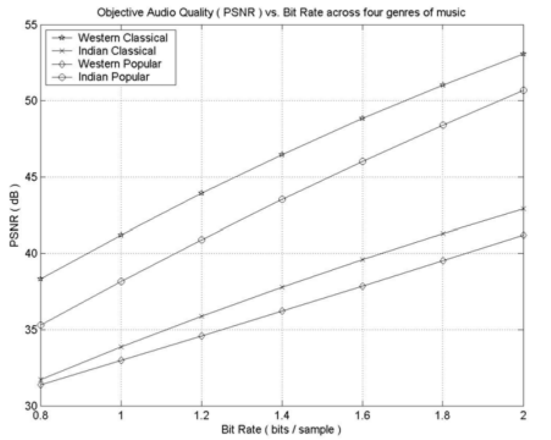

Fig. 6 Graphs of objective audio quality defined by PSNR (dB) as a function of the allowable bit rate (bits / sample) for four different genres. *The varying availability of different genres at different PSNRs and bit rates is easily identifiable, leading to the the proposed scheme of resource dependent genre recommendation*

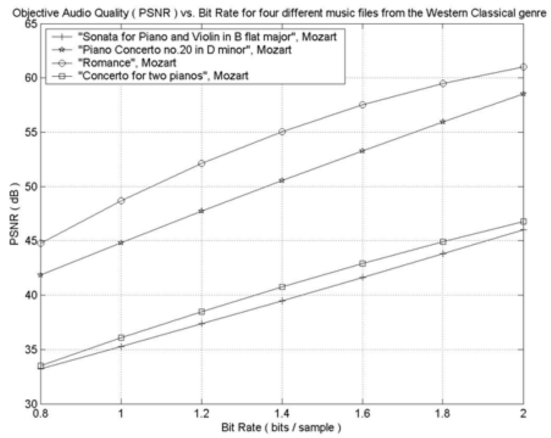

Fig. 7 Graphs of objective audio quality defined by PSNR (dB) as a function of the allowable bit rate (bits / sample) for four songs from Western Classical genre. *The varying availability of different genres at different PSNRs and bit rates is easily identifiable, leading to the the proposed scheme of resource dependent song recommendation*

together are better able to counter this noise.

B] The algorithm goes back to Step 1, now taking the signal $x_1(n)$ in place of $x_0(n)$, and all the sub-steps in 1 are repeated as before, and an approximation, $x_{q,1}(n)$ of $x_1(n)$, and vectors $\mathbf{e}_{q,2}$ and $\mathbf{c}_{q,2}$, are obtained in the same way as before, reducing the noise even further. This loop is to be continued, with the new residue signal at iteration i defined as,

$$x_i(n) = x_{i-1}(n) - x_{q,i-1}(n) \qquad (13)$$

until one of the following cases occur :-

i) The loop has been traversed M times, leading to M

TABLE I
SAMPLE COMPRESSION PERFORMANCE OF THE PROPOSED ALGORITHM OVER A DATABASE OF SIXTEEN SONGS SPANNING FOUR GENRES. UNCOMPRESSED RESOLUTION 16 BITS/SAMPLE, SAMPLING RATE 44.1 KHZ. CLIP LENGTH N = 16000 SAMPLES

Music file	Genre	Compressed bit rate (bits / sample)	PSNR (dB)
Sonata for Piano and Violin in B flat major,Mozart	Western Classical	1.5933	41.8113
Piano Concerto No. 20 in D Minor, Mozart	Western Classical	1.4838	51.6629
Romance,Mozart	Western Classical	1.5129	56.4161
Concerto for two pianos,Mozart	Western Classical	1.5126	42.2866
Raga Mishra Ghara,Ravi Shankar	Indian Classical	1.4966	42.1878
Discovery of India,Ravi Shankar	Indian Classical	1.5036	31.7783
Tala Tabla Tarang,Ravi Shankar	Indian Classical	1.4958	40.9644
Bihag,Ravi Shankar	Indian Classical	1.5826	41.1234
All You Want,Dido	Western Popular	1.5042	37.6118
Thank You,Dido	Western Popular	1.5006	37.1855
Here with me,Dido	Western Popular	1.4594	37.8656
This Land is Mine,Dido	Western Popular	1.4851	36.1252
Aaguner Parashmoni,Hemanta	Indian Popular	1.5081	46.9613
Aaj Khela Bhangar Khela,Hemanta	Indian Popular	1.5151	46.7937
Aaji Marmardhwani Kena,Hemanta	Indian Popular	1.5276	42.8948
Aami Tomay Jato, Hemanta	Indian Popular	1.5061	46.2062

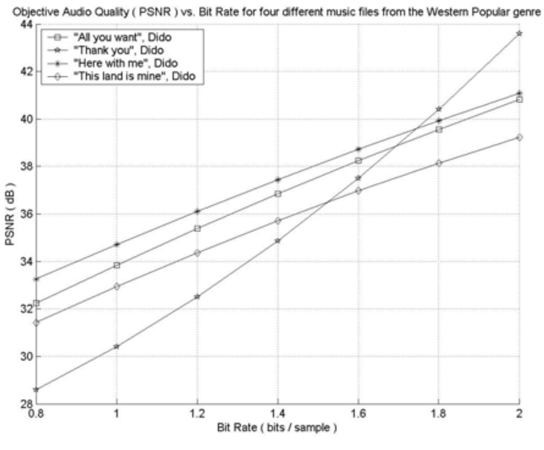

Fig. 8 Graphs of objective audio quality defined by PSNR (dB) as a function of the allowable bit rate (bits / sample) for four songs from Western Popular genre. *The varying availability of different genres at different PSNRs and bit rates is easily identifiable, leading to the the proposed scheme of resource dependent song recommendation*

projection vectors $\{\mathbf{c}_{q,i}\}_{i=1}^{M}$ and M eigenvectors $\{\mathbf{e}_{q,i}\}_{i=1}^{M}$

OR

ii) The reduction in noise from one iteration to the next falls below a preset threshold *th*. This means that the quantization noise has increased so much that even the best eigen-directions are not able to reduce it by a sufficient amount.

If any of the above 2 cases occur, further decomposition into eigen-directions is stopped. If case ii) was true, this actually leads to a saving in the eigenvector overhead, since all of them do not need to be sent now. So, the algorithm goes to Step 3, in order to better utilise the saved amount. If case i) was true, the algorithm is stopped here, having obtained the M coefficient vectors and M eigenvectors required for reconstruction of the signal. [This does not usually occur if the threshold *th* is set accordingly.]

Step 3

A] Since eigen-decomposition is not able to reduce noise sufficiently any more, the remaining residue is directly quantized using the bits left from R after the preceding steps, using Max-Lloyd Algorithm. Next, it is entropy coded. The remaining noise is subtracted from the signal $x_0(n)$ to get the intermediate reconstructed signal $x_{0,int}(n)$.

B] The exact amount of eigenvector overload M_{ovd} that was saved in Step 2 is calculated. This is equal to the number of bands remaining (no. of iterations left before M iterations) times M.

C] An M_{ovd}-length vector $\mathbf{w}$ is calculated which will optimally reconstruct the signal $x_0(n)$ from $x_{0,int}(n)$ by minimizing the Error Criterion $J_\eta(\mathbf{w})$ defined in Section II-E. Here, $L = M_{ovd}$ and $x_\eta = x_{0,int}$.

D] The signal $x_0(n)$ is optimally reconstructed from $x_{0,int}(n)$ using the M_{ovd}-length vector calculated in Step 3C to give $x_{0,rec}(n)$.

An illustrative flowchart for the proposed algorithm is presented in Fig. 5 for clarity.

G. Experimental Evaluation

1) Description of Database and Experimental Setup : For our experiments, a collection of sixteen audio files were chosen from four diverse genres of music, namely a) Western Classical Music (4 files), b) Indian Classical Music (4 files), c) Western Popular Music (4 files) and d) Indian Popular Music (4 files). In each case, the sampling frequency was 44.1 *kHz* and the resolution of the raw audio was 16 bits/sample. For simplicity, only one audio channel (mono) was considered in our experiments.

2) Parameters for Performance Evaluation: For the evaluation of our compression scheme, two established metrics have been considered, namely, the Peak Signal-to-Noise Ratio (PSNR) measured in dB and the compressed Bit Rate in bits/sample [16].

The PSNR is defined as follows,

$$PSNR = 10\log_{10}[\frac{N \max_n x_0(n)^2}{\sum_{n=1}^{N} \eta(n)^2}](dB) \qquad (14)$$

where $\{x_0(n)\}_{n=1}^{N}$ is the audio clip to be compressed while $\{\eta(n)\}_{n=1}^{N}$ is the reconstruction noise defined as,

$$\eta(n) = x_0(n) - x_{0,rec}(n) \qquad (15)$$

where $x_{0,rec}$ is the decompressed of reconstructed signal. The PSNR was calculated after compressing the signal $x_0(n)$ at different bit rates. For our application, the resolution of the raw uncompressed audio data was 16 bits/sample.

3) Performance Results: In Table I, sample PSNR/ bitrate values for all the sixteen audio files in the database have been enlisted. From the table, it can be observed that a reasonably high PSNR value in the range of 35 *dB* to 45 *dB* can be achieved by our proposed compression scheme at bit rates of around 1.5 *bits/sample* over a varied collection of music genres. Such high PSNRs effectively translate to a reconstruction noise well below the perceptual threshold.

Further, at 44.1 *kHz*, a bit rate of 1.5 *bits/sample* is equivalent to 66.15 *kbps* which compares well with current standards [17], [18], [20], [21]. Hence, the QK compression algorithm is efficient as well as effective. A more general perspective is presented in Fig.6, which plots the average zonal objective quality measured in terms of PSNR *versus* bit rates over the four different genres of music considered. The compressed bit rates vary from 0.8 to 2.0 bits/sample at a uniform increment of 0.2 bits/sample.

By observing this graph, it can be noted that different genres of music give strikingly different objective quality at same bit rates. This serves as the guiding philosophy of the AMR system based on the QK compression scheme which is one of the three proposed applications presented in this paper. Even within a particular genre, different songs exhibit major shifts in objective quality at the same bit rate. This is shown Figs. 7 and 8 for the genres of Western Classical and Western Popular music respectively.

III. Proposed Framework for Automatic Music Recommendation

The huge amount of digital music currently available makes it difficult for the average listener to choose his song in an effective way. This has led to the development of automatic music recommendation systems [1] [4] [5]. Current music recommendation systems base their decisions by analyzing past user choices and creating a model for the user or by clustering similar pieces of music together [3]. By contrast, it is known that the choice of a listener is often decided by triple factors, namely, the *bandwidth* available, the *quality* available at that bandwidth and the *genre* preferred by the listener. As established in the previous section, our compression scheme clearly shows how *different* genres exhibit *unequal* PSNRs, and hence disparate audio quality at the same bit rate. Based on this observation, a novel Automatic Music Recommendation system has been proposed here, which could be used separately or in addition to past user choice and is based on the current user constraint viz. bitrate, which is determined by the bandwidth, and the current user preference viz. the quality preferred by the user.

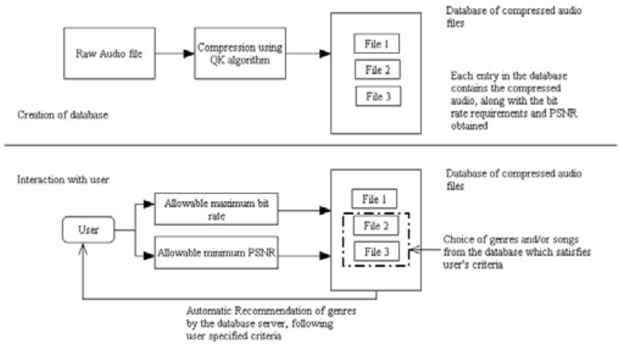

Fig. 5 Schematic block diagram of the proposed Automatic Genre Recommendation scheme

A. Detailed Description of the Proposed AMR Scheme

In the Database Creation stage, new songs are compressed using the proposed QK audio compression algorithm and are stored in the database according to genre, objective quality (PSNR) and bitrate. In the User Interaction stage, the listener or user gives a query for available music genres in the database providing the maximum bitrate that is available to the user, based on the bandwidth. Further, the user provides an objective quality measure in terms of PSNR, indicating the worst audio quality he is willing to accept. The system searches for genres in the database satisfying these two criteria and provides him with a list of choices. As a next step, the user or listener can give a query within his chosen genre for the available songs in the database satisfying his criteria. A comprehensive block diagram is provided in Fig. 5 to explain the main aspects of the system.

B. Sample Interaction with the Database

In Table II, III and IV, *example queries* have been shown, consisting of bitrate and PSNR values and the *search results* based on our Automatic Music Recommendation scheme and the database of sixteen songs previously mentioned. The three tables correspond to the three Objective Quality vs. bitrate graphs shown in Figs. 2,3 and 4 respectively. The first table gives an example of genre recommendation while the next two provide examples of song recommendation.

C. Merits of the Proposed AMR Scheme

Here, the user or listener is provided with a well-defined list of genres and/or songs based on his listening expectations and available bandwidth or bitrate. Should the user wish, he can only be shown genres or songs within a particular genre which can be transmitted with audio quality above a certain threshold, set by the user himself. Hence, the recommendation

TABLE II
SAMPLE STUDY OF GENRE CHOICES AVAILABLE TO A SERVICE USER IN AN AUTOMATIC MUSIC RECOMMENDATION SYSTEM AS A FUNCTION OF AVAILABLE BIT RATE (IN BITS/SAMPLE) AND ACCEPTABLE AUDIO QUALITY (PSNR). BASED ON A DATABASE OF SIXTEEN SONGS FROM FOUR GENRES, VIZ. WESTERN CLASSICAL (CW), INDIAN CLASSICAL (CI), WESTERN POPULAR (PW) AND INDIAN POPULAR (PI). (N.A. - NO AVAILABLE GENRE)

Available bit rate	Minimum acceptable audio quality demanded by user, in terms of PSNR (dB)				
	30	35	40	45	50
0.8	*CW CI PW PI*	*CW PI*	n.a.	n.a	n.a
1.2	*CW CI PW PI*	*CW CI PI*	*CW PI*	n.a	n.a
1.6	*CW CI PW PI*	*CW CI PW PI*	*CW PI*	*CW PI*	n.a
2.0	*CW CI PW PI*	*CW CI PW PI*	*CW CI PW PI*	*CW PI*	*CW PI*

is meaningful with respect to the user. Further, enhanced listening experience is expected to result in increased customer base and higher profits for the database service provider.

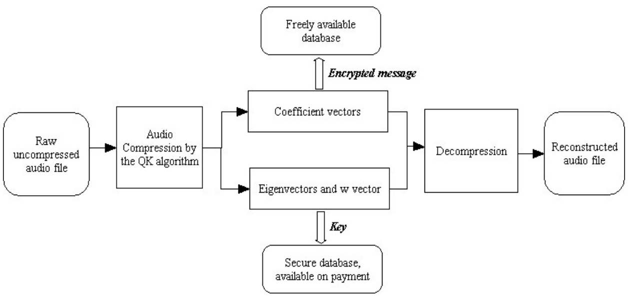

Fig. 6 Schematic block diagram of the proposed Copyright Protection and Digital Rights Management scheme

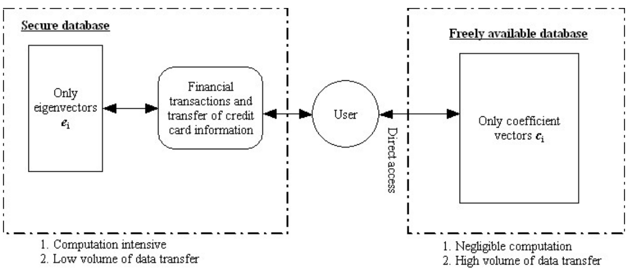

Fig. 7 Detailed block diagram of the Audio Database system incorporating the proposed Copyright protection scheme

TABLE III
SAMPLE STUDY OF SONG CHOICES WITHIN THE WESTERN CLASSICAL GENRE AVAILABLE TO A SERVICE USER IN AN AUTOMATIC MUSIC RECOMMENDATION SYSTEM AS A FUNCTION OF THE AVAILABLE BIT RATE (IN BITS/SAMPLE) AND ACCEPTABLE AUDIO QUALITY (PSNR). CHOICES ARE (A)"SONATA FOR PIANO AND VIOLIN IN B FLAT MAJOR", MOZART, (B)"CONCERTO FOR TWO PIANOS", MOZART, (C)"PIANO CONCERTO NO.20 IN D MINOR", MOZART, (D)"ROMANCE", MOZART. N.A. - NO AVAILABLE SONG

Available bit rate	Minimum acceptable audio quality demanded by user, in terms of PSNR (dB)				
	30	35	40	45	50
0.8	*A B C D*	*C D*	*C D.*	*n.a*	*n.a*
1.2	*A B C D*	*A B C D*	*C D*	*C D*	*D*
1.6	*A B C D*	*A B C D*	*A B C D*	*C D*	*C D*
2.0	*A B C D*	*A B C D*	*A B C D*	*A B C D*	*C D*

TABLE IV
SAMPLE STUDY OF SONG CHOICES WITHIN THE WESTERN POPULAR GENRE AVAILABLE TO A SERVICE USER IN AN AUTOMATIC MUSIC RECOMMENDATION SYSTEM AS A FUNCTION OF THE AVAILABLE BIT RATE (IN BITS/SAMPLE) AND ACCEPTABLE AUDIO QUALITY (PSNR). CHOICES ARE (A) "THANK YOU", DIDO, (B) "THIS LAND IS MINE", DIDO, (C) "ALL YOU WANT", DIDO, (D)"HERE WITH ME", DIDO. N.A. - NO AVAILABLE SONG

Available bit rate	Minimum acceptable audio quality demanded by user, in terms of PSNR (dB)				
	30	35	40	45	50
0.8	*A B C D*	*B C D*	*C D.*	*n.a*	*n.a*
1.2	*A B C D*	*A B C D*	*A B C D*	*B C D*	*D*
1.6	*A B C D*	*A B C D*	*A B C D*	*A B C D*	*A B C D*
2.0	*A B C D*	*A B C D*	*A B C D*	*A B C D*	*A B C D*

IV. PROPOSED FRAMEWORK FOR A DIGITAL RIGHTS MANAGEMENT SYSTEM

A. Description

With the advent of huge databases and ease of copying digital media has grown the risk of copyright insecurity [1]. This has led to the development of digital rights management and copyright protection systems [6]-[8]. Here, an audio database system has been proposed, which incorporates DRM by storing songs after compressing them using the proposed QK Algorithm. As per this algorithm, compression converts an audio file into a set of coefficient vectors and eigenvectors. In this database system, which views the audio database in a cryptographic context, the bulk of the data, ie, the coefficient vectors are considered as the *encrypted message* which are stored in a free database and hence are freely available to the public, while the relatively smaller sized eigenvectors are considered as the *key*. The keys for a particular song are stored in a secure server or database and they are made available only upon payment. Since the key is extracted from the song itself, it is unique. Further, it will be shown that, *if the key or eigenvector of one audio file is used to decompress another audio file, the reconstructed audio compares very poorly against the original file.* Hence, although the bulk of the database is freely available, it cannot be used to duplicate songs because without their proper eigenvector or key, their musical value is lost.

Two detailed block diagrams indicating the major aspects of this database system incorporating digital rights management and copyright protection has been provided in Figs. 6 and 7.

B. Merits of the Proposed Audio Database System

The system is secure. A clever illegal user might try to generate a fake key by applying the QK algorithm on a song which he already possesses. He might choose this surrogate song to be near in genre or style to the song he aims to acquire illegally. However, if he tries to decompress this song using the key of another song, the objective audio quality of the resulting audio will be far below listening standards. This characteristic, clearly shown in Table V, is one of the significant merits of this scheme.

In this scheme, the free database server housing the bulk of the data need not concern itself with payment or security issues. No transactions are required at this end. All the transactions and transfer of credit card information by the user or listener is carried out in the secure server. However, the secure server does not need to concern itself with a large bulk of data since the size of key is very small compared to the size of corresponding song data. For each song, the amount of data to be stored in the secure server, as a percentage of the total amount of audio data to be stored in the free database server can be precisely calculated. The QK audio compression algorithm works by segmenting the raw audio into blocks of equal length, say N. For each block, it calculates a set of M eigenvectors, each of which is defined by M values, for a certain number of levels, say L. Then the size of the key generated, which is nothing but the set of eigenvectors, is $M \times M \times L$ for a block of size N. Hence, the the size of the key is M^2L/N times the size of the entire song data. Using experimentally verified optimal values for the parameters, $N = 16000$, $M = 8$ and $L = 2$, it is found that a mere 0.8% of the total audio data has to be stored as key in the secure server.

Hence, this scheme divides the total workload of a secure music database into processing tasks and data storage and transfer tasks. It is expected that this division would lead to faster overall service time and enhanced listener satisfaction. This is another important merit of this scheme.

It is to be observed that in case of MPEG, there is no concept of data dependent transform and hence the idea of a message related key, ie, eigenvector does not arise. This is an advantageous proposition in our audio compression algorithm over MPEG, in addition to its compression performance.

TABLE V
COMPARISON OF OBJECTIVE AUDIO QUALITY IN TERMS OF PSNR, OBTAINED BY DECOMPRESSING A SONG USING THE CORRECT KEY AND THE HIGHEST POSSIBLE PSNR OBTAINED USING A FAKE KEY GENERATED BY AN ILLEGAL USER BY APPLYING THE QK ALGORITHM ON SOME OTHER SONG. THE COMPARISON IS BASED ON A DATABASE OF SIXTEEN SONGS DRAWN FROM FOUR GENRES, NAMELY WESTERN CLASSICAL (CW), INDIAN CLASSICAL (CI), WESTERN POPULAR (PW) AND INDIAN POPULAR (PI)

Genre and Song no.	CW1	CW2	CW3	CW4	CI1	CI2	CI3	CI4
Correct Key	**319**	**320**	**315**	**310**	**318**	**324**	**310**	**319**
Fake Key	53	58	58	57	42	41	36	51
Genre and Song no.	PW1	PW2	PW3	PW4	PI1	PI2	PI3	PI4
Correct Key	**320**	**315**	**321**	**318**	**316**	**316**	**321**	**317**
Fake Key	45	56	50	29	29	18	45	31

Further, a pricing scheme depending on the number of eigenvectors made available at the secure database can be adopted whereby more eigenvectors shall mean a better audio quality and enhanced listening experience.

IV. PROPOSED FRAMEWORK FOR AUDIO FINGERPRINTING AND DUPLICATE DETECTION

Audio fingerprinting is one of the major issues in Music Information Retrieval research since with the multitude of songs available, proper identification of an unknown audio file has become very cumbersome [1]. Audio fingerprinting is the technique of encapsulating the essence of an audio file into a small fingerprint which is several orders of magnitude smaller in size than the original file [9], [11]. Unknown audio files are identified by comparing their extracted fingerprints with the fingerprints of known files already stored in the database. Possible fakes or duplicates can also be identified by this technique, leading to effective copyright protection [10].

In our scheme, the *eigenvectors* serve as the fingerprint for the song from which they had been extracted through the QK audio compression algorithm. During the Database Creation stage, audio files are compressed using the QK algorithm, their coefficient vectors (which comprise the bulk of the data) are discarded and the eigenvectors are stored along with meta-data information about the audio files. In the Song Identification stage, an unknown audio test input is compressed using QK. In the crucial next step, the compressed audio is reconstructed *not using its own eigenvectors but the fingerprint eigenvectors* stored in the database. PSNR is calculated upon reconstruction by each fingerprint in the database. The idea is that when a song or most likely its noisy version is reconstructed using its own eigenvectors, it will be the best approximation to itself.

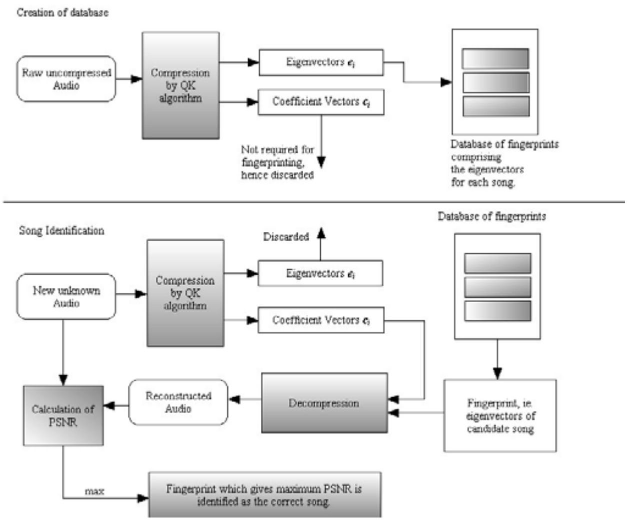

Fig. 8 Schematic block diagram of the proposed Audio Fingerprinting scheme

Using any other eigenvector will lead to a tangible dip in PSNR. Hence, the decision policy adopted is that the *fingerprint giving highest PSNR is linked to the correct identity* of the song.

TABLE VI
TABLE OF PSNR FOR PERFORMANCE EVALUATION OF THE PROPOSED AUDIO FINGERPRINTING SCHEME. THE CORRECT SONG IS IDENTIFIED BY THE FINGERPRINT LEADING TO THE HIGHEST PSNR ON DECOMPRESSION. N.B. T: TEST AUDIO, F: FINGERPRINT. IT CAN BE OBSERVED THAT ALL SIXTEEN THE SONGS IN THE DATABASE CAN BE CORRECTLY IDENTIFIED BY THE PROPOSED SCHEME

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}	F_{15}	F_{16}
T_1	319	34	34	34	42	38	19	53	34	34	43	33	27	19	34	39
T_2	36	320	54	58	36	35	22	36	51	53	36	42	33	22	48	36
T_3	32	50	315	58	32	32	19	32	43	56	32	36	28	19	47	32
T_4	30	52	57	310	30	29	17	30	43	51	30	36	27	17	44	30
T_5	31	29	29	29	318	36	19	31	29	29	36	27	25	19	29	42
T_6	35	33	33	33	41	324	25	35	33	33	40	32	30	26	33	41
T_7	26	26	27	26	26	27	310	27	26	27	27	26	22	36	27	26
T_8	51	29	29	29	34	33	19	319	29	29	35	29	25	19	29	33
T_9	26	47	44	45	26	26	17	26	320	42	26	37	30	17	41	26
T_{10}	36	47	56	52	36	34	18	36	40	315	36	35	26	18	47	34
T_{11}	49	39	40	40	46	40	20	50	37	40	321	33	25	20	40	38
T_{12}	23	29	29	29	22	22	15	23	29	28	22	318	31	14	29	22
T_{13}	16	22	22	22	15	15	9	16	22	22	15	29	316	8	22	15
T_{14}	16	15	15	15	18	17	17	16	15	15	17	14	14	316	15	18
T_{15}	33	44	45	45	33	33	24	33	43	44	34	38	32	24	321	33
T_{16}	34	27	27	27	41	34	14	33	27	27	38	26	24	14	26	317

The proposed scheme has been tested on the database of sixteen files, spanning four diverse genres. As we present in Table VI, the highest PSNR corresponds in each case to the correctly identified song establishing the strength of our fingerprinting technique. Further, the fingerprint is compact and many orders of magnitude smaller than the original file since a set of typically $M^2 = 64$ eigenvectors need to be stored as fingerprint for an audio segment of length $N = 16000$. Finally, it also points out another advantage in using the QK algorithm in the context of MIR.

IV. CONCLUSION

Current audio compression techniques have given rise to a rapid explosion of available digital music. Hence the problem of organization, search and control of such large databases has arisen. In this work, we have shown how a novel compression scheme, the QK algorithm, can be utilized to create an effective Automatic Music Recommendation system, helping the average listener make a meaningful choice of genre and song based on the necessary criteria. Further, a novel digital rights management system is developed which facilitates fast and efficient distribution of digital music content in a secure way, by balancing the workloads of data transfer and information processing. Finally, a new audio fingerprinting technique is proposed based on the QK compression algorithm. Results show that it performs well on a limited database of audio files. Future work will concentrate on further development of each of the three systems introduced.

REFERENCES

[1] Rainer Typke, Frans Wiering, Remco C. Veltkamp, “A Survey of Music Information Retrieval Systems,” *International Symposium on Music Information Retrieval*, 2006.
[2] A. Roy, G. Saha, “Compression using Joint Optimization based on Signal Statistics and Quantization Noise”, *Elesevier Computers and Electrical Engineering.* Submitted for publication.
[3] Shankar Vembu and Stephan Baumann, “A Self-Organizing Map Based Knowledge Discovery for Music Recommendation Systems”, *Lecture notes in Computer Science - Computer Music Modeling and Retrieva*l, Springer Berlin /Heidelberg, vol. 3310/2005, pp. 119 -129.
[4] François Pachet and Pierre Roy, “Automatic Generation of Music Programs”, *Lecture notes in Computer Science - Principles and Practice of Constraint Programming CP99*, Springer Berline / Heidelberg, vol. 1713/2004, pp. 331 - 345.
[5] François Pachet and J.-J.Aucouturier, “Scaling up music playlist generation”, *Proceedings of IEEE International Conference on Multimedia and Expo, 2002*, vol. 1, pp. 105- 108.
[6] Qiong Liu, Reihaneh Safavi-Naini and Nicholas Paul Sheppard, “Digital rights management for content distribution” *Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003*, vol. 21, pp. 49-58..
[7] Frank Hartung and Friedhelm Ramme, “Digital rights management and watermarking of multimedia content for M-commerce applications”, *IEEE Communication Magazine*, vol. 38, no. 11, Nov. 2000, pp. 78-84.
[8] Jong Won Seok and Jin Woo Hong, “Audio watermarking for copyright protection of digital audio data”, *IEEE Electronics Letters*, vol. 37, no. 1, Jan. 2001, pp. 60-61.
[9] P. Cano et al., “Audio Fingerprinting: Concepts And Applications”, *Studies in Computational Intelligence (SCI) 2*, Springer-Verlag 2005, pp.233-245.
[10] Christopher J.C. Burges, Daniel Plastina, John C. Platt, Erin Renshaw, and Henrique S. Malvar, “Using Audio Fingerprinting for duplicate detection and thumbnail generation”, *International Conference on Audio, Speech and Signal Processing, 2005*.
[11] Jason Freeman, “Fast Generation of Audio Signatures to Describe iTunes Libraries”, *Journal of New Music Research 2006*, vol. 35, no. 1, pp. 51-61.
[12] Michail K. Tsatsanis and Georgios B. Giannakis, “Principal Component Filter Banks for Optimal Multiresolution Analysis”, *IEEE Trans. on Signal Proc*., vol. 43, no.8, Aug.1995.
[13] P.Nasiopoulos, M.Yedlin and R.K.Ward, “A high performance fixed-length compression method using the Karhunen-Loeve transform”, *IEEE Trans. on Consumer Elec.*, vol. 41, no. 4, Nov. 1995, pp. 1189-1196.

[14] Anil K. Jain, "A Fast Karhunen Loeve Transform for a class of Random Processes", *IEEE Trans. on Communications*, vol. 24, no. 9, Sept. 1976, pp. 1023-1029.
[15] M. Loève, "Probability Theory - I", New York, Springer-Verlag, 1963.
[16] W. Kinsner, "Compression and Its Metrics for Multimedia", *Proceedings of the First IEEE International Conference on Cognitive Informatics (ICCI02)*, pp. 1-15.
[17] Ahmed H. Tewfik, Deepen Sinha, and Paul Jorgensen, "On the Optimal Choice of a Wavelet for Signal Representation", *IEEE Trans. on Info. Theory*, vol. 38, no. 2, March 1992.
[18] Deepen Sinha and Ahmed H. Tewfik, "Low Bit Rate Transparent Audio Compression using Adapted Wavelets", *IEEE Trans. on Signal Proc.*, vol. 41, no. 12, Dec. 1993.
[19] Jeff B. Burl, "Estimating the Basis Functions of the Karhunen-Loève Transform", *IEEE Trans. on ASSP*, vol. 37, no. 1, Jan. 1989.
[20] D. Pan, "A tutorial on MPEG/audio compression", *Multimedia, IEEE*, vol. 2, no. 2, Summer 1995.
[21] David Salomon, "Data Compression - The Complete Reference", Springer-Verlag, Third Edition, 2004.
[22] Masahiro Nakagawa and Makoto Miyahara, "Generalized Karhunen-Loeve Transformation I (Theoretical Consideration)", *IEEE Trans. on Communications*, vol. COM-35, no. 2, Feb. 1987.
[23] Yingbo Hua and Wanquan Liu, "Generalized Karhunen - Loeve Transform", *IEEE Signal Proc. Letters*, vol. 5, no. 6, June 1998.
[24] Gregory W. Wornell, "A Karhunen-Loève like Expansion for 1/f Proceses via Wavelets", IEEE Trans. on Information Theory, vol.36, no.4, July 1990, pp.859-861.
[25] N.J.Jayant and P.Noll, "Digital Coding of Waveforms", Prentice Hall, Engleood Clffs, NJ, 1984.
[26] Bryan E. Usevitch, "A Tutorial on Modern Lossy Wavelet Image Compression: Foundations of JPEG 2000", *IEEE Signal Proc. Magazine*, Sept. 2001, pp.22 - 35.
[27] Yukihiko Yamashita and Hidemitsu Ogawa, "Relative Karhunen-Lohe Transform", *IEEE Trans. on Signal Proc.*, vol. 44, no. 2, Feb.1996.
[28] Louis L.Scharf and John K. Thomas, "Wiener Filters in Canonical Coordinates for Transform Coding, Filtering and Quantizing", *IEEE Trans. on Signal Proc.*, vol. 46, no. 3, March 1998.

Anindya Roy passed B.E in Electronics and Telecommunication Engineering with highest honours from Bengal Engineering and Science University, Shibpur, Howrah, India in 2005. Presently he is a seniour student of Master of Technology (M. Tech) with Automation and Computer Vision as specialization in the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India. His research areas include audio compression, data compression, music information retrieval, pattern recognition, neural networks, etc.

Goutam Saha graduated in 1990 from Dept. of Electronics & Electrical Communication Engineering, Indian Institute of Technology (IIT), Kharagpur, India. The author worked in Tata Steel, India in the period 1990-1994, joined IIT Kharagpur as CSIR research Fellow in 1994 and completed Ph. D work in 1999. He worked in Institute of Engineering & Management, Salt Lake, Kolkata as a faculty member during 1999-2002 and since 2002 serving IIT Kharagpur as Assistant Professor till date. An active researcher in the field of speech processing, biomedical signal processing, modeling and prediction he has published papers in reputed journals like Physical Review E, IEEE Trans. on Systems, Man & Cybernetics, IEEE Trans. on Biomedical Engineering etc.