A system for determining an identity of a received work. The system receives audio data for an unknown work. The audio data is divided into segments. The system generates a signature of the unknown work from each of the segments. Reduced dimension signatures are then generated at least a portion of the signatures. The reduced dimension signatures are then compared to reduced dimensions signatures of known works that are stored in a database. A list of candidates of known works is generated from the comparison. The signatures of the unknown works are then compared to the signatures of the known works in the list of candidates. The unknown work is then identified as the known work having signatures matching within a threshold.


Business Wire, “Cisco and Fox Host Groundbreaking Screening of Titan A.E.; Animated Epic Will Be First Film Ever to be Digitally Transmitted Over the Internet Monday,” Jun. 5, 2000, 08:14 EDT.

Chen, et al., Yong-Cong, A Secure and Robust Digital Watermarking Technique by the Block Cipher RC6 and Secure Hash Algorithm, Department of Computer Science, National Tsing Hua University, 2001.

Gasaway Laura, Close of Century Sees New Copyright Amendments, Mar. 2000, Information Outlook, 4, 3, 42.

Harris, Lesley Ellen, “To register or not,” Mar. 2006, Information Outlook, 10, 3, 32(a).


* cited by examiner
FIG. 1B

Begin

Receive sampled work

Reduce dimensionality

Determine initial candidates

Search for best candidate

End

104 106 108 110
Audio Signal

Power Spectrum

Mel filters

Segment n

<table>
<thead>
<tr>
<th>MFCC</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.457867</td>
</tr>
<tr>
<td>2</td>
<td>0.098465</td>
</tr>
<tr>
<td>3</td>
<td>-12.2329</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
</tr>
</tbody>
</table>

FIG. 5
Begin  
Create list of candidates  
Perform alternate encoding  
End  

701  
703  

FIG. 7A
Begin

Receive sampled work

Determine initial candidates

Determine best candidate

End

700

702

704

FIG. 7B
METHOD AND APPARATUS FOR IDENTIFYING AN UNKNOWN WORK

PRIORITY CLAIM

This application is a continuation of U.S. patent application Ser. No. 11/239,543, filed Sep. 28, 2005 now U.S. Pat. No. 7,529,659, which is a continuation of U.S. patent application Ser. No. 10/192,783, filed Jul. 9, 2002 now U.S. Pat. No. 6,968,337, which claims the benefit of U.S. Provisional Patent application Ser. No. 60/304,647, filed Jul. 10, 2001, which are hereby incorporated by reference.

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention relates to data communications. In particular, the present invention relates to a novel method and apparatus for identifying an unknown work.

THE PRIOR ART

Background

Digital audio technology has greatly changed the landscape of music and entertainment. Rapid increases in computing power coupled with decreases in cost have made it possible for individuals to generate finished products having a quality once available only in a major studio. One consequence of modern technology is that legacy media storage standards, such as reel-to-reel tapes, are being rapidly replaced by digital storage media, such as the Digital Versatile Disk (DVD), and Digital Audio Tape (DAT). Additionally, with higher capacity hard drives standard on most personal computers, home users may now store digital files such as audio or video tracks on their home computers.

Furthermore, the Internet has generated much excitement, particularly among those who see the Internet as an opportunity to develop new avenues for artistic expression and communication. The Internet has become a virtual gallery, where artists may post their works on a Web page. Once posted, the works may be viewed by anyone having access to the Internet.

One application of the Internet that has received considerable attention is the ability to transmit recorded music over the Internet. Once music has been digitally encoded, the audio may be both downloaded by users for play, or broadcast ("streamed") over the Internet. When audio is streamed, it may be listened to by Internet users in a manner much like traditional radio stations.

Given the widespread use of digital media, digital audio files, or digital video files containing audio information, may need to be identified. The need for identification of digital files may arise in a variety of situations. For example, an artist may wish to verify royalty payments or generate their own Arbitron®-like ratings by identifying how often their works are being streamed or downloaded. Additionally, users may wish to identify a particular work. The prior art has made efforts to create methods for identifying digital audio works.

However, systems of the prior art suffer from certain disadvantages. One area of difficulty arises when a large number of reference signatures must be compared to an unknown audio recording.

The simplest method for comparing an incoming audio signature (which could be from a file on the Internet, a recording of a radio or Internet radio broadcast, a recording from a cell phone, etc.) to a database of reference signatures for the purpose of identification is to simply compare the incoming signature to every element of the database. However, since it may not be known where the reference signatures might have occurred inside the incoming signature, this comparison must be done at many time locations within the incoming signature. Each individual signature-to-signature comparison at each point in time may also be done in a "brute-force" manner using techniques known in the art, essentially computing the full Euclidean distance between the entire signatures' feature vectors. A match can then be declared when one of these comparisons yields a score or distance that is above or below some threshold, respectively.

However, when an audio signature or fingerprint contains a large number of features such a brute-force search becomes too expensive computationally for real-world databases which typically have several hundred thousand to several million signatures.

Many researchers have worked on methods for multi-dimensional indexing, although the greatest effort has gone into geographical (2-dimensional) or spatial (3-dimensional) data. Typically, all of these methods order the elements of the database based on their proximity to each other.

For example, the elements of the database can be clustered into hyperspheres or hyper-rectangles, or the space can be organized into a tree form by using partitioning planes. However, when the number of dimensions is large (on the order of 15 or more), it can be shown mathematically that more-or-less uniformly distributed points in the space all become approximately equidistant from each other. Thus, it becomes impossible to cluster the data in a meaningful way, and comparisons can become both lengthy and inaccurate.

Hence, there exists a need to provide a means for data comparison which overcomes the disadvantages of the prior art.

BRIEF DESCRIPTION OF THE INVENTION

A method and apparatus for identifying an unknown work is disclosed. In one aspect a method may include the acts of providing a reference database having a reduced dimensionality containing signatures of sampled works; receiving a sampled work; producing a signature from the work; and reducing the dimensionality of the signature.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1A is a flowchart of a method according to the present invention.
FIG. 1B is a flowchart of another method according to the present invention.
FIG. 2 is a diagram of a system suitable for use with the present invention.
FIG. 3 is a diagram of segmenting according to the present invention.
FIG. 4 is a detailed diagram of segmenting according to the present invention showing hop size.
FIG. 5 is a graphical flowchart showing the creating of a segment feature vector according to the present invention.
FIG. 6 is a diagram of a signature according to the present invention.
FIG. 7A is a flowchart of a method for preparing a reference database according to the present invention.
FIG. 7B is a flowchart of method for identifying an unknown work according to the present invention.
DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Persons of ordinary skill in the art will realize that the following description of the present invention is illustrative only and not in any way limiting. Other embodiments of the invention will readily suggest themselves to such skilled persons having the benefit of this disclosure.

It is contemplated that the present invention may be embodied in various computer and machine-readable data structures. Furthermore, it is contemplated that data structures embodying the present invention will be transmitted across computer and machine-readable media, and through communications systems by use of standard protocols such as those used to enable the Internet and other computer networking standards.

The invention further relates to machine-readable media on which are stored embodiments of the present invention. It is contemplated that any media suitable for storing instructions related to the present invention is within the scope of the present invention. By way of example, such media may take the form of magnetic, optical, or semiconductor media.

The present invention may be described through the use of flowcharts. Often, a single instance of an embodiment of the present invention will be shown. As is appreciated by those of ordinary skill in the art, however, the protocols, processes, and procedures described herein may be repeated continuously or as often as necessary to satisfy the needs described herein. Accordingly, the representation of the present invention through the use of flowcharts should not be used to limit the scope of the present invention.

The present invention may also be described through the use of web pages in which embodiments of the present invention may be viewed and manipulated. It is contemplated that such web pages may be programmed with web page creation programs using languages standard in the art such as HTML or XML. It is also contemplated that the web pages described herein may be viewed and manipulated with web browsers running on operating systems standard in the art, such as the Microsoft Windows® and Macintosh® versions of the Internet Explorer® and Netscape®. Furthermore, it is contemplated that the functions performed by the various web pages described herein may be implemented through the use of standard programming languages such as Java® and similar languages.

The present invention will first be described in general overview. Then, each element will be described in further detail below.

Referring now to FIG. 1A, a flowchart is shown which provides a general overview of the present invention as related to the preparation of a database of reference signatures. Two overall acts are performed to prepare a reference database in accordance with the present invention: in act 100, the present invention reduces the dimensionality of reference signatures; and the reference database is indexed in act 102.

Referring now to FIG. 1B, a flowchart is shown which provides a general overview of the present invention as related to the identification of an unknown signature in accordance with the present invention. In act 104, a sampled work is received. In act 106, the present invention reduces the dimensionality of the received work. In act 108, the present invention determines initial candidates. In act 110, the present invention searches for the best candidate.

Prior to presenting a detailed overview of each act of FIGS. 1A and 1B, some background will first be presented.

Structural Embodiment of the Present Invention

Referring now to FIG. 2, a diagram of a system suitable for use with the present invention is shown. FIG. 2 includes a client system 200. It is contemplated that client system 200 may comprise a personal computer 202 including hardware and software standard in the art to run an operating system such as Microsoft Windows®, MAC OS® Palm OS, UNIX, or other operating systems standard in the art. Client system 200 may further include a database 204 for storing and retrieving embodiments of the present invention. It is contemplated that database 204 may comprise hardware and software standard in the art and may be operatively coupled to PC 202. Database 204 may also be used to store and retrieve the works and segments utilized by the present invention.

Client system 200 may further include an audio/video (A/V) input device 208. A/V device 208 is operatively coupled to PC 202 and is configured to provide works to the present invention which may be stored in traditional audio or video formats. It is contemplated that A/V device 208 may comprise hardware and software standard in the art to receive and sample audio works (including video containing audio information), and provide the sampled works to the present invention as digital audio files. Typically, the A/V input device 208 would supply raw audio samples in a format such as 16-bit stereo PCM format. A/V input device 208 provides an example of means for receiving a sampled work.

It is contemplated that sampled works may be obtained over the Internet, also. Typically, streaming media over the Internet is provided by a provider, such as provider 218 of FIG. 2. Provider 218 includes a streaming application server 220, configured to retrieve works from database 222 and stream the works in a formats standard in the art, such as Real®, Windows Media®, or QuickTime®. The server then provides the streaming works to a web server 224, which then provides the streamed work to the Internet 214 through a gateway 216. Internet 214 may be any packet-based network standard in the art, such as IP, Frame Relay, or ATM.

To reach the provider 218, the present invention may utilize a cable or DSL head end 212 standard in the art, which is coupled to a cable modem or DSL modem 210 which is in turn coupled to the system's network 206. The network 206 may be any network standard in the art, such as a LAN provided by a PC 202 configured to run software standard in the art.

It is contemplated that the sampled work received by system 200 may contain audio information from a variety of sources known in the art, including, without limitation, radio, the audio portion of a television broadcast, Internet radio, the audio portion of an Internet video program or channel, streaming audio from a network audio server, audio delivered to personal digital assistants over cellular or wireless communication systems, and audio delivered to personal digital assistants over cellular or wireless communication systems.

Additionally, it is contemplated that the present invention may be configured to receive and compare segments coming from a variety of sources either stored or in real-time. For example, it is contemplated that the present invention may compare a real-time streaming work coming from streaming server 218 or A/V device 208 with a reference segment stored in database 204.

Segmenting Background

It is contemplated that a wide variety of sampled works may be utilized in the present invention. However, the inventors have found the present invention especially useful with segmented works. An overview of a segmented work will now be provided.

FIG. 3 shows a diagram showing the segmenting of a work according to the present invention. FIG. 3 includes audio information 300 displayed along a time axis 302. FIG. 3
FIG. 4 shows a detailed diagram of the segmentation of unknown audio according to the present invention. FIG. 4 includes unknown audio information 400 displayed along a time axis 402. FIG. 4 further includes segments 404 and 406 taken of audio information 400 over some segment length T. In an exemplary non-limiting embodiment of the present invention, the segment length of unknown audio segments is chosen to range from 0.5 to 3 seconds.

As can be seen by inspection of FIG. 4, the hop size of unknown audio segments is chosen to be smaller than that of reference segments. In an exemplary non-limiting embodiment of the present invention, the hop size of unknown audio segments is less than 50% of the segment size. In yet another exemplary non-limiting embodiment of the present invention, the unknown audio hop size may be 0.1 seconds.

The inventors have found such a small hop size advantageous for the following reasons. The primary purpose of generating unknown audio segments is to analyze and compare them with the reference segments in the database to look for matches. The inventors have found at least two major reasons why an unknown audio recording would not match its counterpart in the database. One is that the broadcast channel does not produce a perfect copy of the original. For example, the work may be edited or processed or the announcer may talk over part of the work. The other reason is that larger segment boundaries may not line up in time with the original segment boundaries of the target recordings.

The inventors have found that by choosing a smaller hop size, some of the segments will ultimately have time boundaries that line up with the original segments, notwithstanding the problems listed above. The segments that line up with a "clean" segment of the work may then be used to make an accurate comparison while those that do not line up may be ignored. The inventors have found that a hop size of 0.1 seconds seems to be the maximum that would solve this time shifting problem.

As mentioned above, once a work has been segmented, the individual segments are then analyzed to produce a segment feature vector. FIG. 5 is a diagram showing an overview of how the segment feature vectors may be created using the methods described in U.S. Pat. No. 5,918,223 to Blum, et al. It is contemplated that a variety of analysis methods may be useful in the present invention, and many different features may be used to make up the feature vector. The inventors have found that the pitch, brightness, bandwidth, and loudness features of the '223 patent to be useful in the present invention. Additionally, spectral features may be used analyzed, such as the energy in various spectral bands. The inventors have found that the cepstral features (MFCCs) are very robust (more invariant) given the distortion typically introduced during broadcast, such as EQ, multi-band compression/limiting, and audio data compression techniques such as MP3 encoding/decoding, etc.

In act 500, the audio segment is sampled to produce a segment. In act 502, the sampled segment is then analyzed using Fourier Transform techniques to transform the signal into the frequency domain. In act 504, mel frequency filters are applied to the transformed signal to extract the significant audible characteristics of the spectrum. In act 506, a Discrete Cosine Transform is applied which converts the signal into mel frequency cepstral coefficients (MFCCs). Finally, in act 508, the MFCCs are then averaged over a predetermined period. In an exemplary non-limiting embodiment of the present invention, this period is approximately one second. Additionally, other characteristics may be computed at this time, such as brightness or loudness. A segment feature vector
is then produced which contains a list containing at least the 10 MFCC's corresponding average.

The disclosure of FIGS. 3, 4, and 5 provide examples of means for creating a signature of a sampled work having a segment size and a hop size.

FIG. 6 is a diagram showing a complete signature 600 according to the present invention. Signature 600 includes a plurality of segment feature vectors 1 through n generated as shown and described above. Signature 600 may also include an identification portion containing a unique ID. It is contemplated that the identification portion may contain a unique identifier provided by the RIAA (Recording Industry Association of America) or some other audio authority or cataloging agency. The identification portion may also contain information such as the UPC (Universal Product Code) of the various products that contain the audio corresponding to this signature. Additionally, it is contemplated that the signature 600 may also contain information pertaining to the characteristics of the file itself, such as the hop size, segment size, number of segments, etc., which may be useful for storing and indexing.

Signature 600 may then be stored in a database and used for comparisons.

The following computer code in the C programming language provides an example of a database structure in memory according to the present invention:

```c
typedef struct
{
    float hopSize;         /* hop size */
    float segmentSize;     /* segment size */
    MFSignature* signatures;  /* array of signatures */
} MFDatabase;
```

The following provides an example of the structure of a segment according to the present invention:

```c
typedef struct
{
    char* id;           /* unique ID for this audio clip */
    long numSegments;  /* number of segments */
    float* features;   /* feature array */
    long size;         /* size of per-segment feature vector */
    float hopSize;
    float segmentSize;
} MFSignature;
```

The discussion of FIG. 6 provides an example of means for storing segments and signatures according to the present invention.

A more detailed description of the operation of the present invention will now be provided.

Referring now to FIG. 7A, a flowchart showing one aspect of a method according to the present invention is presented. Reference Database Preparation

Prior to the identification of an unknown sample, a database of reference signatures is prepared in accordance with the present invention.

In an exemplary non-limiting embodiment of the present invention, a reference signature may comprise an audio signature derived from a segmentation of the original audio work as described above. In a presently preferred embodiment, reference signatures have 20 non-overlapping segments, where each segment is one second in duration, with one-second spacing from center to center, as described above.

Each of these segments is represented by 10 Mel filtered cepstral coefficients (MFCC's), resulting in a feature vector of 200 dimensions. Since indexing a vector space of this dimensionality is not practical, the number of dimensions used for the initial search for possible candidates is reduced according to the present invention.

Reducing the Dimensionality

FIG. 7A is a flowchart of dimension reduction according to the present invention. The number of dimensions used for the initial search for possible candidates is reduced, resulting in what the inventors refer to as a subspace. By having the present invention search a subspace at the outset, the efficiency of the search may be greatly increased.

Referring now to FIG. 7A, the present invention accomplishes two tasks to develop this subspace: (1) the present invention uses less than the total number of segments in the reference signatures in act 701; and (2) the present invention performs a principal components analysis to reduce the dimensionality in act 703.

Using Less Segments to Perform an Initial Search

The inventors empirically have found that using data from two consecutive segments (i.e., a two-second portion of the signature) to search for approximately 500 candidates is a good tradeoff between computation complexity and accuracy. The number of candidates can be altered for different applications where either speed or accuracy is more or less important.

For example, the present invention may be configured to extract a predetermined percentage of candidates. In an exemplary non-limiting embodiment of the present invention, a list of candidates may comprise 2% of the size of the reference signature database when using 2 segments for the initial search. In another exemplary non-limiting embodiment of the present invention, a list of candidates may be those reference signatures whose distances based on the initial 2 segment search are below a certain threshold.

As will be appreciated by those of ordinary skill in the art, the dimension reduction of the present invention may be used to perform initial search using fewer segments for data other than MFCC-based feature vectors. It is contemplated that any feature-based vector set may be used in the present invention.

Furthermore, the segments used in the initial search do not have to be the same size as the segments used for the final search. Since it may be better to use as few dimensions as possible in the initial search for candidates, a smaller segment size is advantageous here. The full segment size can then be used in the final search. In an exemplary non-limiting embodiment of the current invention, the initial search may use the higher-order MFCCs (since these are the most robust)—this is a simple way to reduce the dimensionality.

In the next section, we will discuss another, more sophisticated, method for reducing the segment size for the initial candidate search.

Perform Alternate Encoding

The second step is to use an alternate encoding of the MFCC data which has the same information but with fewer features.

To accomplish this, the present invention first performs an eigenanalysis of N candidates to determine the principal components of the MFCC's for our typical audio data. In an exemplary non-limiting embodiment of the present invention, the present invention examines 25,000 audio signatures of 20 segments each—each taken from a different recording, which gives provides 500,000 sets of MFCC's. The inventors have found that this is enough to be a good statistical sample of the feature vectors.
As is appreciated by those of ordinary skill in the art, the number examined in the present invention may be adjusted to provide a good statistical sample of different kinds of music. For example, 100 or a 1000 segments may be satisfactory.

Next, a Karhunen-Loève transformation is derived. Each set of 10 MFCCs becomes a column of a matrix A. We then compute \( A^T A \) and find the 10 eigenvalues and eigenvectors of this matrix. Sorting the eigenvectors by eigenvalue (largest eigenvalue first) results in a list of orthogonal basis vectors that are the principal components of the segment data. For a database of typical music recordings, 95% of the information in the MFCCs is contained in the first 7 components of this new basis.

As is known by those having ordinary skill in the art, the Karhunen-Loève transformation is represented by the matrix that has the all 10 of the above eigenvectors as its rows. This transformation is applied to all the segments of all the reference signatures in the database as well as to all the segments of any signatures that are to be identified. This allows approximate distances to be computed by using the first few components of the transformed segment MFCC vectors for a small tradeoff in accuracy. Most importantly, it reduces the initial search dimension to 14 (7 components times 2 segments), which can be indexed with reasonable efficiency.

As will be appreciated by those of ordinary skill in the art, dimension reduction according to the present invention may be utilized to examine subspaces for feature sets other than MFCCs. The dimension reduction of the present invention may be applied to any set of features since such sets comprise vectors of floating point numbers. For example, given a feature vector comprising spectral coefficients and loudness, one could still apply the KL-process of the present invention to yield a smaller and more easily searched feature vector.

Furthermore, the transform of the present invention may be applied to each segment separately. For example, prior art identification methods may use a single 30-second segment of sound over which they compute an average feature vector. Of course, the accuracy of such methods are much lower, but the process of the present invention may work for such features as well. Moreover, such prior art methods may be used as an initial search.

The dimension reduction aspect of the present invention provides significant efficiency gains over prior art methods. For example, in a "brute force" method, the signature of the incoming sampled work is tested against every reference signature in the database. This is time-consuming because the comparison of any two signatures is a 200-dimensional comparison and because there are a lot of reference signatures in the database. Either alone are not unsatisfactory, but both together takes a long time. The present invention solves the first problem by searching only a subspace, i.e., using less than all 200 dimensions in the comparison.

In addition to the raw speedup given by searching a subspace, the reduced dimensionality also allows one to practically index the database of reference signatures. As mentioned above, it is impractical to index a 200-dimensional database, but 14 is practical.

The present disclosure thus provides for several manners in which the dimensionality may be reduced:

1. searching for the top N candidates over a subspace;
2. searching for the top N candidates using less than the total number of segments from the reference signature;
3. searching for the top N candidates by projecting the reference signatures and signature of the work to be identified onto a subspace, where the subspace is determined by a Karhunen-Loève transformation.

The preparation of the reference database may occur at any time. For example, the results of the preparation may occur each time the server is started up. Additionally, the results could be saved and reused from then on, or the results may be prepared once and used over again. It may need to be recomputed whenever a new reference signature is added to the database.

Computing the Index

The present invention may also compute an index of the reference signatures. As is appreciated by those having ordinary skill in the art, many indexing strategies are available for use in the present invention. Examples include the k-d tree, the SS-tree, the R-tree, the SR-tree, and so on. Any look-up method known in the art may be used in the present disclosure. Common to all indexing strategies is that the multidimensional space is broken into a hierarchy of regions which are then structured into a tree. As one progress down the tree during the search process, the regions become smaller and have fewer elements. All of these trees have tradeoffs that affect the performance under different conditions, e.g., whether the entire tree fits into memory, whether the data is highly clustered, and so on.

In an exemplary non-limiting embodiment of the present invention, a binary k-d tree indexing method is utilized. This is a technique well-known in the art, but a brief overview is given here. At the top level, the method looks to see which dimension has the greatest extent, and generates a hyperplane perpendicular to this dimension that splits the data into two regions at its median. This yields two spaces on either side of the plane. This process is continued by recursion on the data in each of these subspaces until each of the subspaces has a single element.

After the reference database has been prepared, the present invention may be used to identify an unknown work. Such a process will now be shown and described.

Identification of an Unknown Work

Referring now to FIG. 7B, a flowchart of a method for identifying an unknown work is shown. In act 700, the present invention receives a sampled work. In act 702, the present invention determines a set of initial candidates. Finally, in act 704, the present invention determines the best candidate. Each act will now be described in more detail.

Receiving a Sampled Work

Beginning with act 700, a sampled work is provided to the present invention. It is contemplated that the work will be provided to the present invention as a digital audio stream. It should be understood that if the audio is in analog form, it may be digitized in any manner standard in the art.

Indexed Lookup

In act 702, the present invention determines the initial candidates. In a preferred embodiment, the present invention uses the index created above to perform an indexed candidate search.

An index created in accordance with the present invention may be used to do the N nearest neighbor search required to find the initial candidates.

Candidate Search

Once a set of N nearest neighbors is determined, the closest candidate may then be determined in act 704. In an exemplary non-limiting embodiment of the present invention, a brute-force search method may be used to determine which candidate is the closest to the target signature. In another preferred embodiment, the present invention may compare the distance of this best candidate to a predetermined threshold to determine whether there is a match.
There are a number of techniques that may be applied to the candidate search stage which make it much faster. In one aspect, these techniques may be used in a straightforward brute-force search that did not make use of any of the steps previously described above. That is, one could do a brute-force search directly on the reference signature database without going through the index search of step 702, for example. Since there is some overhead in doing step 702, direct brute-force search may be faster for some applications, especially those that need only a small reference database, e.g., generating a playlist for a radio station that plays music from a small set of possibilities.

Speedups of Brute-Force Search.

Any reference signature that is close to the real-time signature has to be reasonably close to it for every segment in the signature. Therefore, in one aspect, several intermediate thresholds are tested as the distance is computed and the computation is exited if any of these thresholds are exceeded. In a further aspect, each single-segment-to-segment distance is computed as the sum of the squared differences of the MFCCs for the two corresponding segments. Given the current computation of the MFCCs, average segment-to-segment distances for matches are about approximately 2.0. In an exemplary non-limiting embodiment of the invention, we exit the computation and set the distance to infinity if any single segment-to-segment distance is greater than 20. In further aspects, the computation is exited if any two segment-to-segment distances are greater than 15, or if any four segment-to-segment distances are greater than 10. It should be clear to anyone skilled in the art that other thresholds for other combinations of intermediate distances could easily be implemented and set using empirical tests.

Since any match will also be close to a match at a small time-offset, we may initially compute the distances at multiples of the hop size. If any of these distances are below a certain threshold, we compute the distances for hops near it. In an exemplary non-limiting embodiment of the invention, we compute distances for every third hop. If the distance is below 8.0, we compute the distances for the neighboring hops. It should be clear to anyone skilled in the art that other thresholds for other hop-skippings could easily be implemented and set using simple empirical tests.

While embodiments and applications of this invention have been shown and described, it would be apparent to those skilled in the art that many more modifications than mentioned above are possible without departing from the inventive concepts herein. For example, the teachings of the present disclosure may be used to identify a variety of sampled works, including, but not limited to, images, video and general time-based media. The invention, therefore, is not to be restricted except in the spirit of the appended claims.

The invention claimed is:

1. A method, implemented by a computing system programmed to perform the following, comprising:
   providing, by the computing system, a reference database containing a plurality of reference works, wherein each reference work corresponds to a reference signature comprising a plurality of reference segments;
   receiving, by the computing system, an unknown work, wherein the unknown work comprises a plurality of unknown segments;
   calculating, by the computing system, at least one distance from one of the plurality of unknown segments to one of the plurality of reference segments;
   if the at least one calculated distance does not meet a predefined threshold, repeating, by the computing system, a calculation for a distance from each subsequent segment of the plurality of unknown segments to each subsequent segment of the plurality of reference segments until the predefined threshold is met; and
   identifying, by the computing system, the unknown work as one of the plurality of reference works based on the calculated distances.
2. The method of claim 1, wherein the predetermined threshold includes one calculated distance greater than 10.
3. The method of claim 1, wherein the predetermined threshold includes two calculated distances greater than 15.
4. The method of claim 1, wherein the predetermined threshold includes four calculated distances greater than 10.
5. The method of claim 1, wherein calculating the at least one distance from one of the plurality of unknown segments to one of the plurality of reference segments comprises:
   calculating a mel frequency cepstral coefficients (MFCC) for the one of the plurality of unknown segments;
   calculating a MFCC for the one of the plurality of reference segments; and
   calculating a sum of squared differences of the MFCCs for the unknown segment and the
   6. A non-transitory computer readable storage medium, comprising executable instructions which when executed on a processing system cause the processing system to perform a method comprising:
   providing a reference database containing a plurality of reference works, wherein each reference work corresponds to a reference signature comprising a plurality of reference segments, the segments defining a hop size;
   receiving an unknown work having an unknown signature;
   calculating a distance from the unknown signature to one of the plurality of reference signatures at a point in time in said reference signature;
   if the calculated distance is greater than a predefined threshold, repeating a calculation for a distance from the unknown signature to the reference signature at least one offset point in time in the reference signature until the calculated distance is less than the predefined threshold;
   calculating a distance from the unknown signature to the reference signature at one or more points in time neighboring the point in time where the distance calculated is less than the predefined threshold; and
   identifying the unknown signature as one of the plurality of reference works based on the distances calculated at the one or more neighboring points in time.
7. The non-transitory computer readable storage medium of claim 6, wherein the offset point in time is a point in time one or more multiples of the hop size from the last point in time where the distance was calculated.
8. The non-transitory computer readable storage medium of claim 6, wherein the offset point in time is a point in time three hop sizes from the last point in time where the distance was calculated.
9. The non-transitory computer readable storage medium of claim 6, wherein the one or more neighboring points in time is one or more point in time one hop size from the point in time in the reference signature where the distance calculated is less than the predefined threshold.
10. The non-transitory computer readable storage medium of claim 6, wherein the predefined threshold is 8.0.
11. A computer-implemented apparatus comprising:
   a memory to store a reference database containing a plurality of reference works, wherein each reference work corresponds to a reference signature comprising a plurality of reference segments;
   a processor coupled to the memory.
to receive an unknown work, wherein the unknown work comprises a plurality of unknown segments; to calculate at least one distance from one of the plurality of unknown segments to one of the plurality of reference segments; if the at least one calculated distance does not meet a predefined threshold, to repeat a calculation for a distance from each subsequent segment of the plurality of unknown segments to each subsequent segment of the plurality of reference segments until the predefined threshold is met; and to identify the unknown work as one of the plurality of reference works based on the calculated distances.

12. The system of claim 11, wherein the predetermined threshold includes one calculated distance greater than 20.

13. The system of claim 11, wherein the predetermined threshold includes two calculated distances greater than 15.

14. The system of claim 11, wherein the predetermined threshold includes four calculated distances greater than 10.

15. The system of claim 11, wherein the processor is to calculate the at least one distance from one of the plurality of unknown segments to one of the plurality of reference segments by:
   calculating a mel frequency cepstral coefficients (MFCC) for the one of the plurality of unknown segments;
   calculating a MFCC for the one of the plurality of reference segments; and
   calculating a sum of squared differences of the MFCCs for the unknown segment and the reference segment.

* * * * *