Feature Extraction Approach to Enhance Information Retrieval for Musical Media

By

Ahmed M. Dalli Al-juffer

Supervisor

Moayad A. Fadhil Al-athami

This Thesis was Submitted in Partial Fulfillment of the Requirements for the Master's Degree in Computer science

Deanship of Academic Research and Graduate Studies

Philadelphia University

May 2013

جامعة فيلادلفيا

نموذج تفويض

انا احمد ماهر دللي افوض جامعة فيلادلفيا بتزويد نسخ من رسالتي للمكتبات والمؤسسات والهيئات التعليمية عند طلبها

التوقيـع:

التاريخ:

Philadelphia University

Authorization Form

I am Ahmed Maher Dalli authorize Philadelphia University to supply copies of my thesis for libraries and educational institutions and organizations upon request

Signature:

Date:

Feature Extraction Approach to Enhance Information Retrieval for Musical Media

By

Ahmed M. Dalli Al-juffer

Supervisor

Moayad A. Fadhil Al-athami

This Thesis was Submitted in Partial Fulfillment of the Requirements for the Master's Degree in Computer science

Deanship of Academic Research and Graduate Studies

Philadelphia University

May 2013

Successfully defened and approved on ……………………………

Examination Committee		Signature
Academic Rank:	**Chairman**	
Dr. Moayad A. Fadhil Al-athami		……………………………..
Academic Rank:	**Member**	
Dr. Mohammad Mustafa Taye		……………………………..
Academic Rank:	**External Member**	
Dr. Faruq Al-omari		…………………………….

DEDICATION

I like to get this opportunity to donate this project for

My Parents, brothers and sisters,

And friends for their precious support in my life.

May God bless them

ACKNOWLEDGMENT

Words are not enough to convey my deep sense of gratitude & gratefulness to my worth supervisor Dr. Moayad

All the credits of my final year project to all my teachers in computer science department who paid keep interest & very good Attention in my project

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACM-SIGMM	Association of Computing Machinery-Special Interest Group on Multimedia
BOF	Bag-of-features
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
HDP	Hierarchical Dirichlet Process
LPC	Linear Predicative Coding
MFCC	Mel Frequency Cepstrum Coefficients
MIR	Music Information Retrieval
SNRs	Signal to Noise Ratios
ZCR	Zero Crossing Rate

LIST OF SYMBOLS

F_M	13-dimesional Mel-frequency cepstral coefficient (MFCC) vector
F_S	4-dimensional spectroscopic profile vector
F_T	timber structure appearance vector
x_i	input sub-melody features
$d(X,Y)$	Distance
P	distance order
$X = (x_1, x_2, \ldots., x_n$	Database
n	database elements number
$\|\{A\}$	argument (A)
E	energy feature
ZCR	zero-crossing rate
r	Correlation
x and y	Pieces
$\bar{x}$ and $\bar{y}$	means of both pieces x and y
N	Number of samples in the piece
$\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \ldots, \mathrm{x}_\mathrm{n})$	training data
$\mathrm{Y} = (\mathrm{y}_1, \mathrm{y}, \ldots, \mathrm{y}_\mathrm{n})$	testing data

ABSTRACT

In this research, the implementation of advanced music information retrieval system using the MATLAB software program is proposed and analyzed. This implemented system retrieves musical data based on initially pre-processing the musical data using three main techniques; Discrete Wavelet Transform (DWT), Linear Predictive Coding (LPC) and filtering function, after that, extracting seven features; energy in DWT of the high band, Energy in DWT of the low band, total energy of both bands, zero crossing rate of the high band, zero crossing rate of the low band, autocorrelation of the high band and autocorrelation of the low band from these data and then matching the extracted features with those of training musical data that are stored in a predefined database. The matching process depends on computing the Minkowski distance among musical data.

Results demonstrate that the implemented music retrieval system can effectively and accurately retrieve the correct musical data based on the chosen testing data and comparing them with predefined ones.

CHAPTER ONE

INTRODACTION

1.1INTRODUCTION

In comparison with a vocal signal, a music signal is likely to be more stationary and possesses some very specific properties in terms of musical tones, intervals, chords, instruments, melodic lines and rhythms. While many effective and high performance Music Information Retrieval (MIR) algorithms have been proposed, most of these works tend to consider a music signal as a vocal one and make use of Mel Frequency Cepstrum Coefficients (MFCC) based features which are primarily designed for speech signal processing. These coefficients were introduced in the 60’s and used since that time for speech signal processing. The MFCC computation averages spectrum in sub-bands and provide the average spectrum characteristics. Whereas there averages are inclined to capture the global tone of a music signal and claimed to be of use in music information retrieval, they cannot characterize the required music properties; musical tones, intervals, chords, instruments, melodic lines and rhythms as needed for perceptual understanding by human beings and quickly find their limits. Mainly, the existing ceiling can be overcome based on combining the spectral similarity descriptors with high-level analysis (SUNG. et al, 2009 POPE. et al, 2004, LINDENBAUM.et al, 2010).

The objective of this project is to perform music retrieval by similarity and automatic labelling (music genre classification) by introducing an approach to high-level musical analysis.

1.2BACKGROUND

Recently, music is available in digital forms in which personal collections can simply go over the practical restrictions on the time required to listen for these music. Mainly, the traditional ways that are used to listen for music and the used methods for determining music, as the record stores and radio broadcasts are substituted by modified ways, such as websites which offer rapid update for new music determination channels (LAMERE, 2008, CASEY.et al, 2008).

In general, the music users are not the only persons who have prospects of searchable music collections, since with the increase of user activities in digital music, the opportunities to investigate into utilizing large collections of music to detect both patterns and trends in music are increased. Therefore, the used strategies for allowing accessing several music collections must be improved to continue with prospects of both search and browse functionality. These strategies are called Music Information Retrieval (MIR) (CASEY.et al, 2008).

The main three audiences of the MIR strategies are; the incorporated industry bodies in disseminating, recording and aggregating music, the users who need to find music and utilize it in an improved way and several professionals, as music teachers, performers and producers. In Music Information Retrieval (MIR) systems, the multi-disciplinary field focuses on finding out information from music to use them in solving several problems, such as kind classification, beat detection, music recommendation and automatic music transcription (CASEY.et al, 2008).

1.3AIM

This research aims to introduce and implement a robust retrieval and matching technique for processing the musical data and investigating several music databases by contents using the MATLAB software program. It depends on using music features and their corresponding similarity measures which can eliminate the problem of characterizing the music quality properties and improve the Music Information Retrieval (MIR) system performance

1.4OBJECTIVES

This project will be based on music features and corresponding similarity measures for music information retrieval. While popular spectrum related techniques tend to characterize music tone properties, musical features can help to overcome the existing limits and, hence, to enhance the performance of music information retrieval algorithms. The project will also try to extract music properties such as rhythmic, melodic, tonality and timbre fingerprints for automatic music search by similarity as well as automatic classification. Furthermore, the objectives –in general- for this research will include:

- Define the problem of low performance in musical retrieval.
- Analyze and study the existed features of musical extraction.
- Develope or enhance a new methodology in order to be used in the feature extraction for better performance and to overcome the existing current limits.
- Apply the enhancement on the retrieval information for musical media.
- Derive conclusions and future recommendations.

1.5PROBLEM STATEMENT

Generally, several Music Information Retrieval (MIR) algorithms have been introduced and published. These algorithms depend on characterizing the music quality properties in order to extract information from music, which is considered as a real problem. In addition, these algorithms may expose to various problems in detecting beats, recognizing artists, classifying genres and recommending music. Thus, the need for more improved algorithms that can extract the most important musical information in a direct way from the audio based on using the musical features is increased.

The objective of this project is to perform music retrieval by similarity and automatic labelling (music type classification) by introducing an approach to high-level musical analysis.

1.6RESEARCH IMPORTANCE

As a major product for entertainment, there is a huge amount of digital musical content produced, broadcasted, distributed and exchanged. There is a rising demand for content-based music search services. Similarity-based music navigation is becoming crucial for enabling easy access to the ever-growing amount of digital music available to professionals and amateurs alike. A professional user, such as a radio programmer, may want to search for a different interpretation of one song to include in a radio playlist. In addition, a radio programmer needs to discover new songs and artists to help his listeners to discover new music. The music amateur on the other hand has different needs, ranging from active music discovery for the fans, to the simple seed song playlist generation of similar items. Such ways to organize musical collections as genre classification and title structuring are important as they facilitate music navigation and discovery.

The research will offer a robust retrieval and matching technique for processing the musical data and searching for several music databases based on their contents using the MATLAB$^{(TM)}$ software program. It will depend on extracting seven music features; Energy in Discrete Wavelet Transform (DWT) of the high band, Energy in Discrete Wavelet Transform (DWT) of the low band, total energy of both bands, zero crossing rate of the high band, zero crossing rate of the low band, autocorrelation of the high band and autocorrelation of the low band from both the training and testing data and then performing a matching process using the Minkowski distance. This technique will be applied on a database that contains 200 musical data. These data will be processed based on applying three main techniques; Discrete Wavelet Transform (DWT), Linear Predictive Coding (LPC) and filtering function

1.7CONTRIBUTIONS

This project offers an enhancement for the available Music Information Retrieval (MIR) systems based on the following approaches:

- The system accuracy is enhanced based on pre-processing the musical data by applying both the normalization and learning processes
- The problems of the used features are solved based on extracting more helpful features; Energy in DWT of the high band, Energy in DWT of the low band, total energy of both bands, zero crossing rate of the high band, zero crossing rate of the low band, autocorrelation of the high band and autocorrelation of the low band.
- Errors are minimized based on applying other improved techniques; Discrete Wavelet Transform (DWT) and filtering in addition to the use of Linear Predicative Coding (LPC).

1.8PROJECT OUTLINE

This research is divided into five main chapters as follows:

- Chapter one: Introduction

This chapter gives an introduction about this research and demonstrates the main aims, objectives, problem statement and research importance of this project

- Chapter two: Literature review

This chapter reviews some of the related works about the music information retrieval systems.

- Chapter three: System analysis and design

This chapter gives an introduction about the MIR systems and proposed the main architecture of this system, the used algorithms for feature extraction, the extracted statistical features, the normalization process and the learning process

- Chapter four: System implementation

This chapter demonstrates the implementation of the proposed music retrieval system and the matching process

- Chapter five: Conclusion, evaluation and future work

This chapter concludes the proposed work, compares the obtained results in this work with some of the related works and illustrates the main works that can be done in the future to improve the proposed system

CHAPTER TWO

LITERATURE REVIEW

2.1OVERVIEW

In the late of the 20th century, both the space of storage and the processing power turned out to be cheaper than before and the compressed formats of files were improved in order to facilitate the store of large music amounts with high quality. According to Wang et al (2006) and Orio (2006), the proposed proceeds are related to the fact that music still very agreeable yet when it is saved or in other cases transmitted in a digital way, the need for retrieving data from music is increased and the probabilities to concentrate on these needs are created, unlike other forms of art, as painting.

According to Schiiuble (1997), music is considered as a form of art that can cross the languages barriers and various cultural environments. Thus, people with various backgrounds can share the music. Schiiuble explored that the Music Information Retrieval (MIR) is committed in order to meet the music data users' needs. According to Huron (2000), there are several aims of using music; the exercises trainer inquires about a definite tempo, the director of a film inquires about music that assigning a convinced mood, the physiotherapist inquires about music which inspires a specified patient, the driver of a vehicle inquires about music that keeps them attentive and an advertiser inquires about a melody that is extremely unforgettable.

According to Kolakowska et al (2008), various MIR systems depend on the techniques of symbolic matching and some of them depend on the techniques of audio matching, thus these systems use broad musical features range that needs a preprocessing step in order to extract various notes, such as the pitch, intervals and duration. Miotto and Orio (2010) explored that other MIR systems depend on features that have high musical level, as the contour, rhythm and timber, conversely, none of the MIR systems search music utilizes the wavelet transforms which depend on features.

Sung et al (2009) published paper titled "Effective Digital Music Retrieval System through Content-based Features". They proposed effective system for digital music retrieval. They divided the proposed system into Client and Server. Client part consists of pre-processing and Content-based feature extraction stages. In preprocessing stage, they minimized Time code Gap that is occurred among same music contents. As content-based feature, first-order differentiated MFCC were used as a dataset of features. Server part included Music Server and Music Matching stage. Extracted features from 1,000 digital music files were stored in Music Server. In Music Matching stage, they found retrieval result through similarity measure by Dynamic Time Warping (DTW). In experiment, they used 450 queries. These were made by mixing different compression standards and sound qualities from 50 digital music files. Retrieval accuracy indicated 97% and retrieval average time was 15ms in every single query. The experiment proved that proposed system is effective in retrieving digital music and robust at various user environments of web.

Aucouturier et al (2005) discussed the Electronic Music Distribution and how it needs robust and automatically extracted music features. An important attribute of a piece of polyphonic music is what is commonly referred to as “the way it sounds”. While there has been a large quantity of research done to model the timbre of individual instruments, little work has been done to analyze “real world” timbre mixtures such as the ones found in popular music. They presented their research about such “polyphonic timbres” and then they described an effective way to model the textures found in a given music signal, and then they show that such timbre models provide new solutions to many issues traditionally encountered in music signal processing and music information retrieval. Notably, they described their applications for music similarity, segmentation and pattern induction

Tzanetakis and Cook (2002) studied the classifications of audio signal in Musical Genre. During their search they presented the automatic classification of audio signals into a hierarchy of musical genres is explored. More specifically, three feature sets for representing tumbrel texture, rhythmic content and pitch content are proposed. The performance and relative importance of the proposed features is investigated by training statistical pattern recognition classifiers using real-world audio collections. Both whole file and real-time frame-based classification schemes are described. Using the proposed feature sets, classification of 61% for ten musical genres is achieved. This result is comparable to results reported for human musical genre classification.

2.2ADVANCED MULTIMEDIA INFORMATION RETRIEVAL

The multimedia retrieving systems imply the search for different Medias, like pictures, melodies and films. Among the development quantity of melodies, pictures and films on users' computers and through the Internet, the contribution of efficient investigation for the required media is developed rapidly. The main challenge that faces the developers and evaluators of Medias is that these Medias can be represented in several ways (MAYBURY, 1997, STEPHEN, 2003).

Lately, investigators of Music Information Retrieval (MIR) depend on extracting the main substances of music instead of features to retrieve musical data which help in ensuring the users' needs. Users utilize several MIR types through Google AltaVista Auditory and pictures and video explore. Whereas modern arts, the methods passing MIR to the normal clients (LEW. et al, 2006).

A number of workshops and meetings about the MIR were set up. The most famous meetings are the meetings that incorporate the Association of Computing Machinery-Special Interest Group on Multimedia (ACM-SIGMM), plus the Global meetings for pictures and films recovery. This treats the MIR approaches (REN and BRACEWELL, 2009).

Lew et al 2006. Offered a couple of essential demands of the MIR methods: investigation and "browsing and summarizing a media collection", the two systems to completing the demands separate to two groups: sort based on group and the feature based on group. The sort based on group offers the opportunity to media semantically that lead to enhance the recovery which let this group to become more popular (REN and BRACEWELL, 2009, LEW. et al, 2006).

2.3METADATA-BASED AND CONTENT-BASED APPROACHES

Generally, the main two approaches that used for the MIR are the metadata based approaches and content based approaches. The metadata based approaches depend on finding the helpful groups that can be used to describe music, distinguish among various recordings or artists and organize the descriptions types that presented in the music. On the other hand, the content based approaches are used to find the types of descriptions in the music as well as they can be applied on another musical content feature, such as the contents of vocal music (WIERING, 2006).

According to Wiering (2006), the most challenged approaches in the MIR are the approaches that cope with the real musical contents, such as the melody and pitch since the musical contents appear in two different formats; notation and sound where the most usual one is choosing the sound format. Practically, extracting features from different musical signals and distinguishing among melodies, pitches, harmonies, beat patterns and the used instruments are considered as simple functions that can be performed easily by humans. On the other hand, extracting these features and utilizing them by other people to retrieve music is a real problem. In addition, the polyphony dictation which is the music where various pitch events occur at the same time is a real challenge too. Mainly, there are some methods that are independent on the pitch detection, such as the ones introduced by Pickens et al (2003) and Casey (2006).

Several researchers, such as Castan (2006) have used in their investigations the symbolic representations of musical signals based on encoding the musical scores in different obtainable encoding systems.

2.4 APPLICATIONS OF MIR

The main three applications of music information retrieval are: query by example, query by humming, and type classification (KAMINSKAS and RICCI, 2012):

The query by example depends on using audio signals as inputs and then returns several data about the recording, such as title, artist and type. According to Cano et al (2005), this application depends on using the fingerprinting technique that includes two main stages; extraction of features and matching to represent the input musical signal in a unique way based on using its features. This technique is useful for recognizing specified musical signals not all the work.

The query by humming depends on using hummed melodies by users as inputs and returns the matched tracks and their information. According to Birmingham et al (2006), the fingerprinting technique cannot be used in the query by humming application since it aims to distinguish the distorted song versions which cannot be retrieved by the query by humming system. This technique is appropriate for melodic music, not rhythmic compositions.

For the type classification application, huge music collections can be organized based on assigning the type labels for all the music tracks which in turn simplifies the navigation and investigation for music contents. On the other hand, the main challenge in this application is the uncertainty of the type concept. According to Pachet and Cazaly (2000), there is no recognized general classification of the music types.

2.5 LITERATURE REVIEW ABOUT MUSIC

2.5.1 A SURVEY OF AUDIO-BASED MUSIC CLASSIFICATION AND ANNOTATION

A musical tune is a composition of sounds that resulted from various musical instruments, such as piano. It consists of many notes where each note can be defined as a sign used in a music notation system to signify the related pitch and duration of a specific sound. The music notation system signifies the supposed music in an oral way based on using written symbols. Generally, the song feature can be represented as an individual vector that specialized for every melody. The matching determines come back for every couple of melodies, or the characteristic group of a defined characteristic vectors. The illustration of the individual vector is direct closing to use the normal or the medium rate for very characteristic feature in excess for every part in order to create a song-level abstract feature vector. The consequential "feature vector" may be employed forward via a typical organizer for classification. This classification essentially used for temporary features and MFCC features (TZANETAKIS and COOK, 2002, LIDY and RAUBER, 2005, MANDEL and ELLIS, 2005, SEYERLEHNER. et al, 2008).

The individual features vector may be estimated for a melody through utilizing an international "codebook" form. The international "codebook" created through collecting the restricted characteristic vector plus using the collection attention in a "codebook vectors". Every confined vector in a melody will be marked to the closest "codebook vector". Then, the feature will be estimated including the amount of occurrence of the "codebook vectors". The feature vector for the melody may be utilized immediately picked by the charts allocation of the" codebook vectors". In order to form the allocation of the confined "feature vectors" in the codebook forming the expansion of the international "codebook" forms was suggested via utilizing "Hierarchical Dirichlet Process (HDP)"(FOOTE, 2008, HOFFMAN. et al, 2008).

The melody likeness regularly estimated via balancing the temporary tone features in every melody. The huge amount of the accessible features related to the exact numerical forming has been the most important reason for the attention on the timbre features. In order to form the allocation of the features the possibility form was calculated for every song. Via using the calculated possibility forms for pair of melodies depending on the number of differences measurements the variation between them will be obvious (LOGAN and SALOMAN, 2001, AUCOUTURIER and PACHET, 2002, PAMPALK. et al, 2003).

2.5.2 MUSIC CLASSIFICATION

An original form which utilized for melody classification in order to beat the problem faces the "feature illustrations" is the Bag-of-features (BOF). This new form leaves no doubt on essential possibilities allocation. Therefore it may enhance the protecting the important data in the "feature". The ability for discovering of mixing many of codebooks within the BOF forms in order to decrease the variation within the "codebooks construction"(FU. et al, 2011).

2.6FEATURE REPRESENTATION AND EXTRACTION

The appearance for melody types sorting was tested into two sets: timber surface and musical substance appearance. Tone appearances shows the short duration characters, like the spectroscopic data. Musical substance appearance shows long duration character counting the tone, hit, rhythm and substance. The equipments of the rhythm act own a very important affect in identification of the sort. "The timbre" essential voice supplies are reversed on the spectroscopic allocation of a rhythm signs (ULAS. et al. 2007).

The timbre structure appearances that typify to a short duration spectroscopic characters are supposed to show melodic sorts. The short period examination is executed for upon 25 ms extending beyond auditory window in order to draw out the timber structure appearance for all the 10 ms borders. The auditory window implemented for exaggeration window with size 25 ms in order to take away the outcomes that produced from the edges. The timber structure appearance, consisting of Mel-frequency cepstral coefficient (MFCC) vector F_M and dimensional spectroscopic profile vector F_S with spectroscopic centric, spectroscopic turn off, spectroscopic flow and non intersection range, removed away from the examination window. Thus, this formula $F_T = \ [F_M{}'F_S{}']$ can be used to represent the overall mix tone structure appearance vector (ULAS. et al. 2007).

CAPTER THREE
SYSTEM ANALISIS AND DESIGN

3.1IMPLEMENTED MUSIC RETRIEVAL SYSTEM

This research introduces the implementation of a Music Information Retrieval (MIR) system to improve innovative techniques in order to process the musical information and search various databases of music via content. Thus, the need for forceful matching and retrieval techniques increases. The following chart summarizes the implemented Music Information Retrieval (MIR) system.

The proposed system is divided into two phases; offline and online. In the offline phase, the extracted features from known data are extracted and then saved in a database, while in the online phase, features of the unknown data are compared with those of the known data to find a match between them and then retrieve the most matched data. The used database in this work consists of thirteen musical data.

As shown in the figure, the main stages of the implemented Music Information Retrieval (MIR) system are: Inserting a musical data, preprocessing the data based on applying both the normalization and learning process methods, applying three main techniques to extract features; Discrete Wavelet Transform (DWT), Linear Predictive Coding (LPC) and filtering techniques, extracting features, comparing between features based on calculating the Minkowski distance and then retrieving the most matched data.

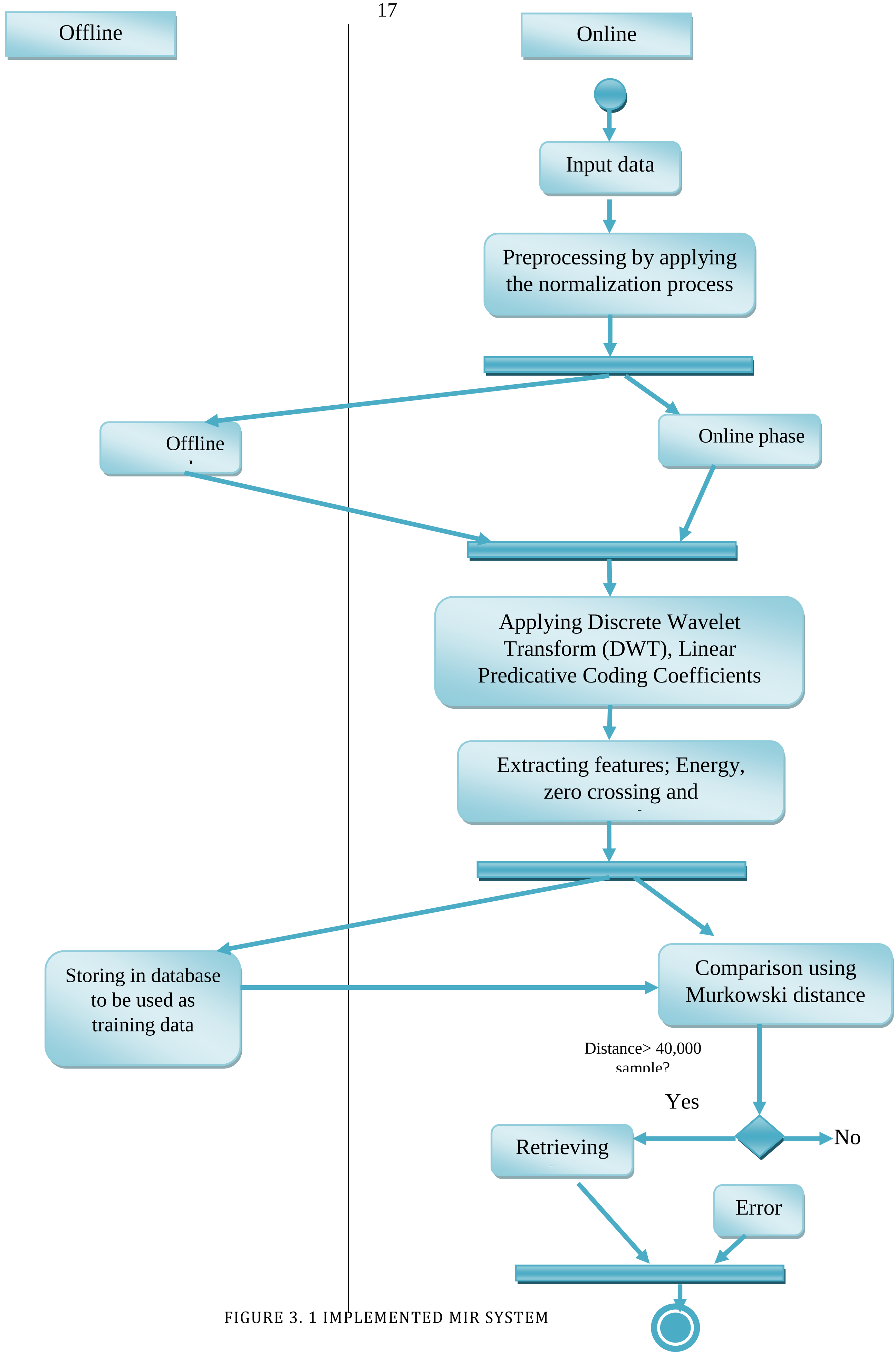

FIGURE 3. 1 IMPLEMENTED MIR SYSTEM

3.2PREPROCESSING METHODS

The used thirteen normal musical data are initially preprocessed based on applying two main techniques; normalization and learning process. These two techniques normalize the used musical signal into the same duration and then divide them into two musical signals; training and testing with the same duration. These two techniques are explored in the following subsections:

3.2.1 NORMALIZATION

The normalization is the process of scanning the required audio files, determining the mean level of amplitude, data or time duration and then decreasing or increasing the files levels in which all the sounds will be at the same volume. This process can be applied on the musical data, amplitudes and time durations. This process is divided into two stages. In the first stage, the most frequently appearing level during the music piece is defined as the temporary fundamental level. Figure 3.2 shows an example of how the most frequently appearing note duration appears in the samples of a musical signal where the horizontal axis represents the percentages of the appearance of the most frequently appearing note duration in a sample of the musical signal, while the vertical axis represents the number of samples in the musical signal. In the proposed figure, the most frequently appearing note duration appears in the range 40%–60% in various samples of the musical signal.

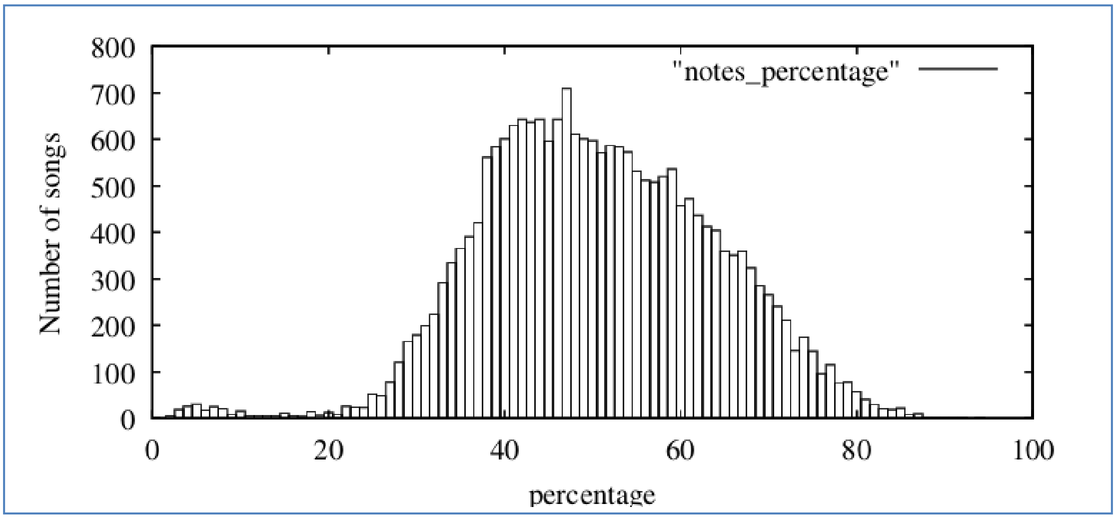

FIGURE 3. 2 THE MOST FREQUENTLY APPEARING NOTE DURATION IN THE SAMPLES OF A MUSICAL SIGNAL

After defining the most frequently appearing note duration, the music data are temporarily normalized by transforming all the musical levels into levels that are relative to the temporary fundamental level. The normalized data are then divided into multi music segments that are called subsequences via utilizing a dividing window method. These subsequences which have the same musical information amount depending on the temporary fundamental level are generated (KOSUGI, 2004).

In the second stage, the most frequently appearing level in each one of the generated subsequences is chosen as the fundamental level for each one of the subsequences. This unit is called the real fundamental level. If a subsequence has different temporary and real fundamental levels, then the subsequence will be re-normalized again depending on the real fundamental level and the musical information amount in this subsequence is changed depending on the real fundamental level. The generated segments after the change of the musical information amount are called the sub-subsequences which are generated by suing a sliding window method in which the lengths of both the slide and window are calculated again depending on the real fundamental level and the utilized for the second dividing process (KOSUGI, 2004).

Figure 3.3 illustrates an example of a time normalization using the proposed two stages in order to generate subsequences and sub-subsequences:

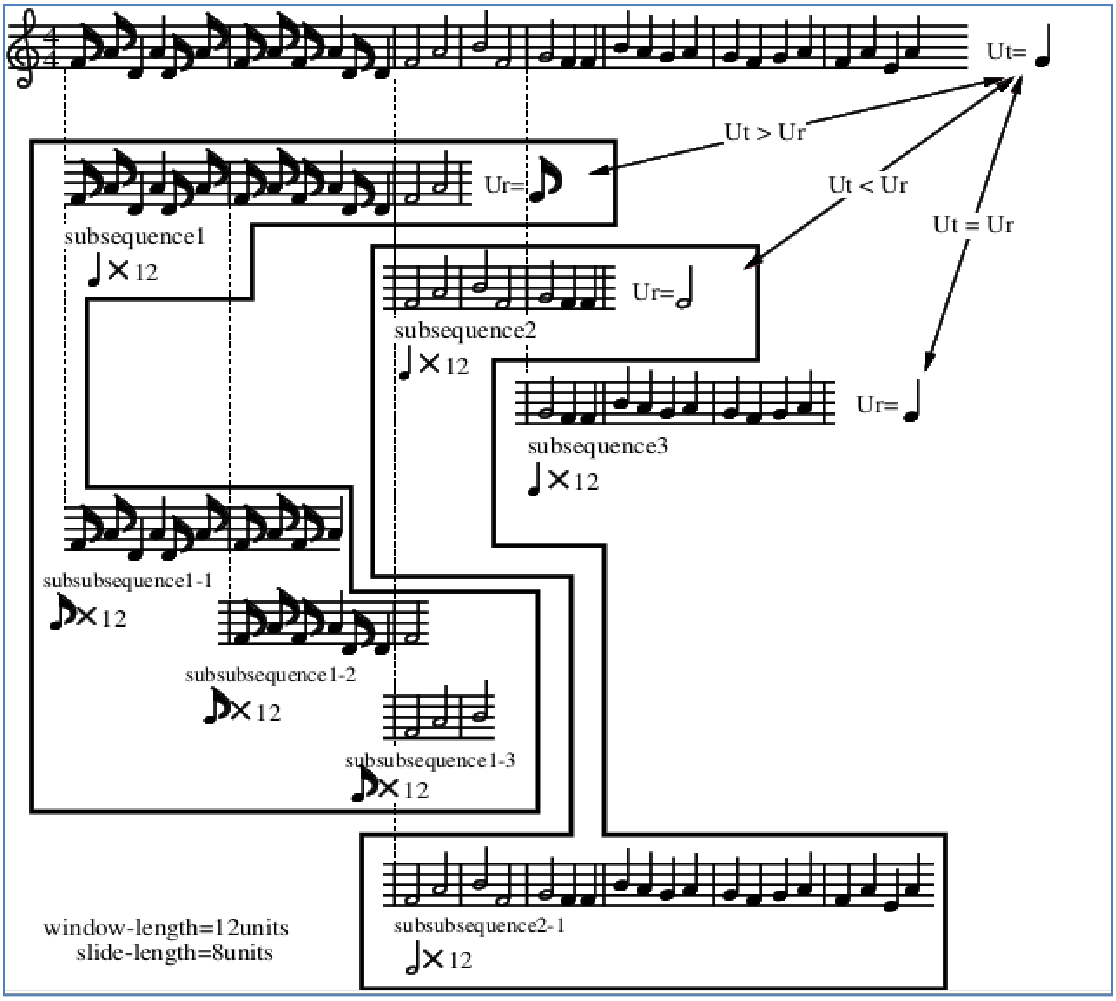

FIGURE 3. 3 AN EXAMPLE OF A TIME NORMALIZATION PROCESS (KOSUGI, 2004)

- In the first stage, the most frequently appearing note duration, which is a quarter note (one beat) is defined and called the temporary basic unit length(U_t).
- In the second stage, a new unit that is called the real basic unit length(U_r) is defined. After that, three subsequences of twelve beats are produced every eight beats. These three generated subsequences are: $U_t > U_r$ (left subsequence), $U_t < U_r$ (right subsequence) and $U_t = U_r$ (middle subsequence).
- The most frequently appearing note duration in each subsequence is selected and set as U_r.
- When there is a difference between both the U_t and U_r in any subsequence, this subsequence is normalized again and the musical data amount is changed

depending on the U_r. Thus, new segments are produced and called sub-subsequences.

- Since there is a difference between both the U_t and U_r in the left and right subsequences, these two subsequences are re-normalized.
- As shown in figure 3.3, for the first subsequence ($U_t > U_r$), the most frequently appearing note duration is an eighth note. Therefore, U_r is an eighth note (half beat). Since there is a difference between both the U_t and U_r, this subsequence is re-normalized into three new sub-subsequences (sub-subsequences 1-1, 1-2 and 1-3) where each one has six beats and between each two sub-subsequences there are four beats.
- For the second subsequence ($U_t < U_r$), the most frequently appearing note duration is also a half note (two beats). Therefore, U_r is a half note. Since there is a difference between both the U_t and U_r, this subsequence is re-normalized one sub-subsequence 2-1with twenty four beats
- For the third subsequence ($U_t = U_r$), subsequence ($U_t < U_r$), the most frequently appearing note duration is a quarter note, thus, both U_r and U_t are a quarter note. Since there is no difference between both the U_t and U_r, no sub-subsequences are produced.

The 200 used music data are normalized into the same duration (20 seconds) to divide each one into two data with the same duration. In addition, the amplitudes of each data are normalized based on dividing it by the maximum amplitude.

3.2.2 LEARNING PROCESS

The learning is the process of analyzing the problem of the function estimation theoretically from a known data collection and creating various practical algorithms to estimate functions. The learning process is applied based on using block similarity features in order to recognize the correct music. Based on this process, the system can recognize and retrieve the same music data with different time slots. In other words, is music A is stored in a database between 2-4 seconds, the system can recognize and retrieve it for various time slots, such as between10-20 seconds.

The main themes of the learning process in the statistics are: prediction, classification and clustering. The prediction theme represents the process of making suggestions from data that are subjected to random variations based on using known datasets. The classification theme represents the categorization of unknown musical samples into known classes using a learned model that is constructed using training data (known data). The clustering theme represents the grouping of unknown data in a way that data in the same group (cluster) are more similar to each other than data in other clusters. These themes can be used in various applications, such as speech understanding, texts classification, computer vision and bioinformatics (LU, 2008).

The learning process aims to offer a frame for analyzing the inference problem which is of having knowledge, making different models based on known datasets and then making predictions. This is analyzed using a statistical learning strategy that contains three stages; data collection, model construction and prediction. In the data collection stage, the required known (training) data are collected and stored in a database to be used to train the system model. In the model construction stage, an effective model is constructed based on the generated dataset. In the prediction stage, the constructed model is tested based on classifying unknown (testing) data into known classes. The following figure summarizes the proposed learning process.

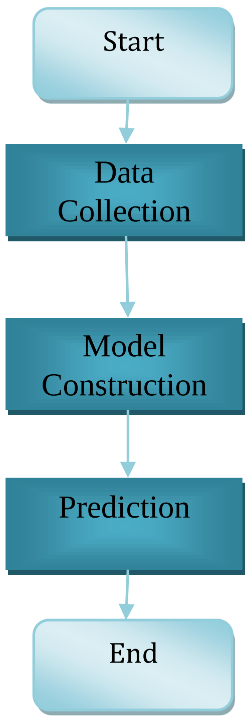

FIGURE 3. 4 THE STRATEGY OF THE LEARNING PROCESS

The data collection stage accounts for an unexpectedly large cost of improving the learning process. A preliminary possibility study can be performed by using a small dataset of standard examples. But, there is a need for more data in order to assure an enhanced performance in the learning system. The main limitations of this stage are: this stage may not be a simple process, the collection of data may be complicated because of the inputs variation and the labeling of large datasets is a time consuming process. This problem can be solved based on collecting data automatically (using computers) instead of collecting them manually.

In the model contraction stage, a robust model is constructed based on the learning from the specified training data. Mainly, the learning method is divided into two parts; the supervised and unsupervised methods. The learning method is called an unsupervised method when there is no prior data are given. Thus, samples are classified into various groups based on using a clustering algorithm that depends on the distance or the similarity measure among samples. The unsupervised learning method has some limitations, such as: the prior data that contains information about which samples are estimated to a specific group together or create a categorizer and utilize it to estimate some unknown samples cannot be used. Therefore, the supervised learning method that is implemented for the categorization and estimation is more commonly utilized.

When training data are given, then the construction of function that matches the required data precisely is always possible. On the other hand, when a noise is presented, then it will not be the best thing to be done and it may give a poor performance. Thus, the main concept in designing learning algorithms is the looking for regularities in the practical phenomenon, such as the training data where this can be indiscriminate from the past to the future.

In the prediction stage, the unknown samples are classified into known groups depending on the learned model. The learned model is used to find out the best solution to a known problem. On the other hand, the learning process in the Multimedia Information Retrieval (MIR) faces three main problems; high dimensionality, small sample size and semantic gap which is the gap between the semantic concepts that are set by users and the image features, such as color and shape which are related to the whole image or a certain region of the image.

3.3TECHNIQUES FOR FEATURES EXTRACTION

Three main techniques are then applied on the processed musical signals in order to extract their features. These techniques are: Discrete Wavelet Transform (DWT), Linear Predicative Coding Coefficients (LPC) and filtering techniques. Table 3.1 below illustrates these techniques. One of the newest algorithms that are used in the MIR is the ℓ^1 minimization, where ℓ^1 is a form of geometry where the metric or main distance function of the Euclidean geometry is substituted by another metric where the distance between two points is the summation of the absolute differences of their main coordinates. This algorithm is faster than the traditional methods which depend on characterizing the music quality properties instead of extracting features since it has the ability to substitute the iterative linear solvers that are used in the computations of the Fourier domain with considerable time savings (GOLDSTEIN and OSHER, 2008).

TABLE 3. 1 THE MAIN USED ALGORITHMS (AIQUTT. ET AL, 2011)

Notation	Description
Discrete Wavelet Transform (DWT)	It is a linear transformation that depends on a vector of data where its length is an integer power of two. This function converts the vector into a numerically different vector with the same length
Linear Predicative Coding Coefficients (LPC)	It offers the forward linear analyst coefficients based on decreasing the calculation error in the least squares sense
Filtering function	It filters a sequence of data depending on a digital filter that works in both real inputs and complex inputs.

As proposed, the main used algorithms in the features extraction are: Linear Predictive Coding (LPC) model and Discrete Wavelet Transform (DWT). The filtration function is used to filter out the work using MATLAB.

3.3.1 LINEAR PREDICTIVE CODING (LPC) TECHNIQUE WITH HAMMING FILTERATION FUNCTION

In the LPC model, the features are divided into two main features groups; frequency domain features, such as Discrete Cosine Transform (DCT) coefficients and time domain features, such as ZCR and energy features.

The LPC can be obtained based on reducing the mean square error among the real samples of the musical data and the predicted ones based on using the autocorrelation method. It depends on applying a windowing technique, such as the Hamming window which is a weighted moving average conversion that is utilized to smooth the squared coefficients of each frequency of the signal. The used windowing techniques prevent the spectral artifacts which are resulted from the discontinuities in the frame endpoints. In addition, the LPC prevents the generation of sound interruptions at the end of the windowing technique. The following figure shows an example of applying the hamming window on a signal.

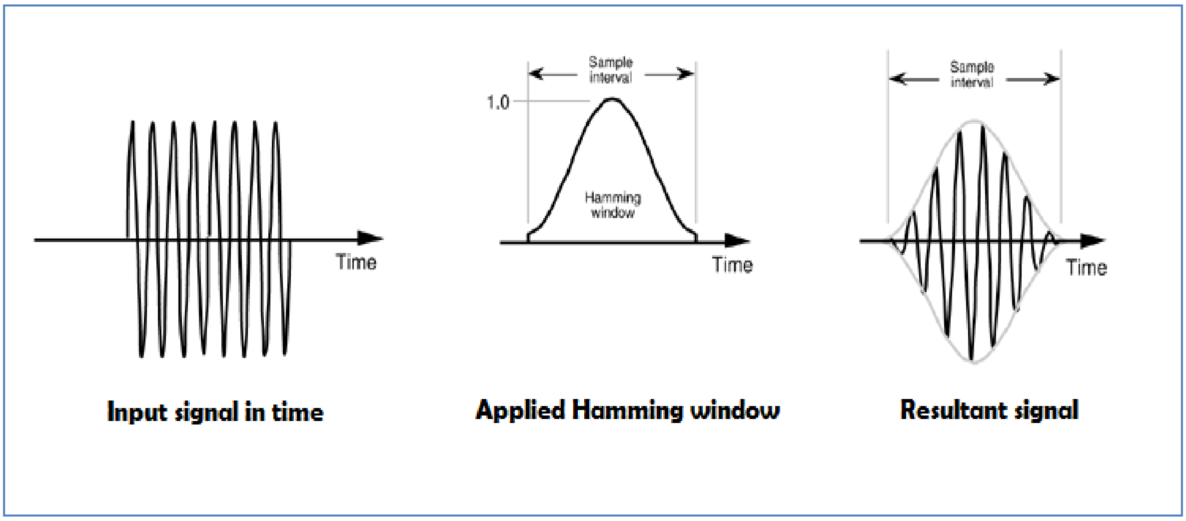

FIGURE 3. 5 EXAMPLE OF APPLYING THE HAMMING WINDOW

The LPC technique is utilized broadly in the signal processing, particularly in the applications of the speech processing. It supposes that the present samples can be estimated via combining its past samples. In addition, this technique can approximate the significant frequencies systems, such as the musical instruments. The LPC method takes signals in frames to verify both the input signal and the filter coefficients. It depends on using two processes; encoding and decoding; the encoding process verifies a precise parameters set that can be used in the modeling, while the decoding process utilizes the obtained parameters from the encoding process to construct a synthesized version from the original signal. The LPC coefficients can be calculated in the time domain (TSAI and WANG, 2006).

The LPC techniques are improved in order to generate frequency domain features since the majority of the musical signals perceptual information is in the frequency domain. Figure 3.6 shows the LPC model. LPC model consists of six main stages; pre-emphasis, frame blocking, windowing, autocorrelation analysis, LPC analysis and LPC parameter. The pre-emphasis is the first stage that is applied to reduce noises in which the weak signal parts and the high frequencies are improved before transmitting or recording them. It offers high Signal to Noise Ratios (SNRs) (MOSA and ALI, 2009).

In the frame blocking stage, the digitalized musical signal is blocked into frames that have N samples where the neighboring frames are separated with M samples. The windowing is used to reduce the discontinuity in order to avoid the signal spectral leakage at both the start and end of all the frames in which each frame is multiplied via a window function. In the autocorrelation analysis, each one of the windowed signal frames is auto-correlated based on the LPC analysis order which is between 8 and 16. In the LPC analysis, each one of the auto- correlated frames is converted into an LPC parameter set of coefficients (MOSA and ALI, 2009).

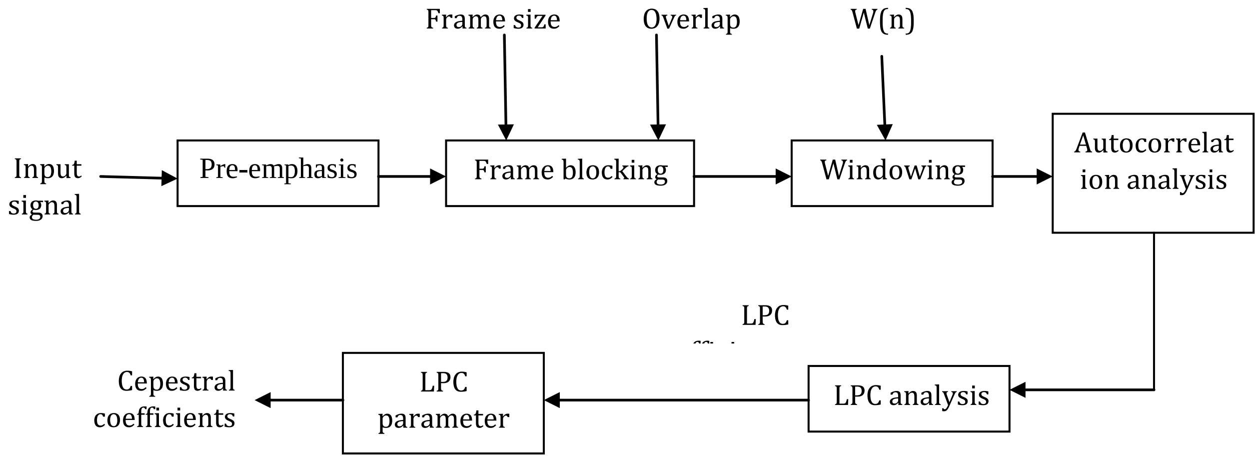

FIGURE 3. 6 THE LPC MODEL

3.3.2 DISCRETE WAVELET TRANSFORM (DWT) TECHNIQUE

The wavelet analysis is a windowing technique that uses regions with variables sizes. It permits the utilization of long-time intervals in which more accurate low frequency data are needed and short-time intervals in which more high frequency data are needed. The main benefits of designing wavelets are offering high-quality time resolution at elevated frequencies and offering high-quality frequency resolution at small frequencies. Property of disappearance moment permits wavelet to focus on significant data and remove noisy signal, while the property of decorrelated coefficients allows wavelet to decrease the sequential correlation, thus the wavelet coefficients correlations are smaller than the ones of the related sequential process. Therefore, after wavelet, the complicated signal which is in the time domain is decreased into a simpler process in the domain of wavelet.

The main used wavelet is the Daubechies wavelet. The first Daubechies wavelet is called Haar wavelet which is the basic and simple Daubechies wavelet as well as it is quick and simple to be concerned on various signals (ALQUTT.et al, 2011).

The main aim of the wavelet is to offer time-scale analysis relative results versus both the spectrum and time domains. The spectrum domain includes both the pitch and FFT, while the time domain includes the statistical features. The Discrete Wavelet Transform (DWT) is a linear transformation that performs on a vector of data in which its length is an integer power of two. This function converts the vector into a numerically different vector with the same length. It is considered as a tool for separating the data into various frequency components and then studying each one of these components with resolution matched to its scale. It can be calculated with a cascade filters.

The wavelet transform represents the summation of all the signal samples multiplied by wavelet function scaled shifted version or as called the mother wavelet (Ψ). The mother wavelet can be selected from basically infinite selections of short and oscillatory signals with no DC offset.

Generally, there are several mother wavelets that can be utilized in the design of eavelet transform technique, such as Haar, Morlet, Symmlet and Daubechies. The selection of the optimal mother wavelet depends on taking into account the attributes of all the available mother wavelets.

3.4EXTRACTED FEATURES

After applying the proposed three techniques on the used musical data, features can be extracted. Suppose that a database is $\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \dots., \mathrm{x}_\mathrm{n})$, where n represents the number of musical data in the database. Two databases are implemented in this project, which are: files and features databases. The files database includes the real .wav files, while the features database includes all the files feature vectors. These databases are implemented in the beginning of utilizing the proposed system where they must be updated after each addition of new files. The real wav files are added to the first database in which their features are extracted and then saved in the second database. Thus, each one of the entries in the second database is related to its corresponding wav file in the first database. Figure 3.7 below shows the proposed process.

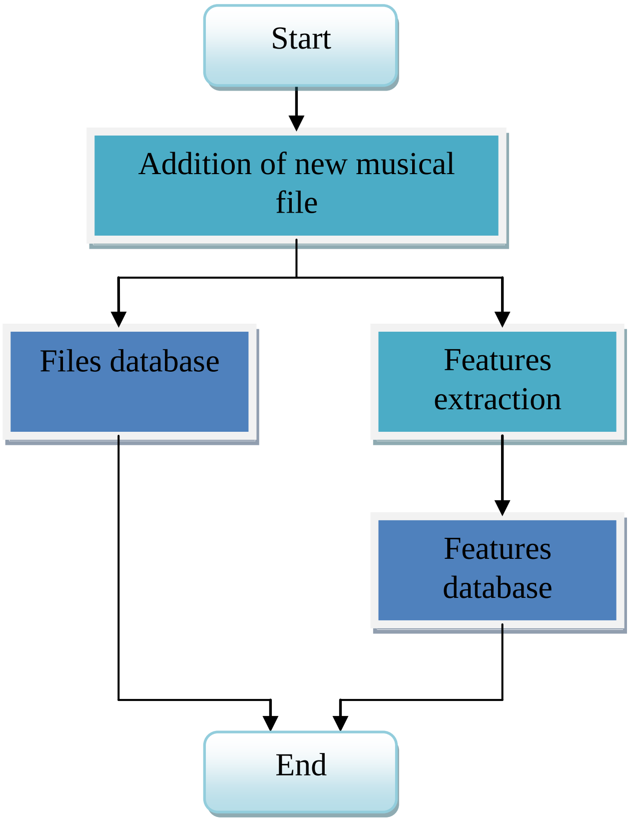

FIGURE 3. 7 ADDITION OF NEW FILES TO THE DATABASES

The used database includes 200 musical data files that are normalized into the same running time duration which equals to 20 seconds. Each one of these data is divided into two data; a training data that represents the first 10 seconds of the data and a testing data which represents the other 10 seconds of the data. The Discrete Wavelet Transform (DWT) is used to convert each one of the 60 data (200 training data and 200 testing data) into two numerically different bands with the same length; high band and low band. The main statistical features that are extracted from each data are illustrated in the following table.

TABLE 3. 2 THE MAIN STATISTICAL FEATURES

Features	Description
Energy in DWT of the high frequency band	It reproduces the signal amplitude differences and offers foundations for both unvoiced and voiced pieces that have high frequencies

Energy in DWT of the low frequency band	It reproduces the signal amplitude differences and offers foundations for both unvoiced and voiced pieces that have low frequencies
Total energy of both bands	It is the summation of the proposed two energies in DWT
Zero crossing (changes of a musical signal from positive to negative or vice versa) of the high frequency band	It is the summation of the crossing per each frame (axis) of the high frequency band
Zero crossing (changes of a musical signal from positive to negative or vice versa) of the low frequency band	It is the summation of the crossing per each frame (axis) of the low frequency band
Autocorrelation of the high frequency band	It represents the similarity among samples as a function of time for the high frequency band
Autocorrelation of the low frequency band	It represents the similarity among samples as a function of time the low frequency band

The energy feature reproduces the signal amplitude differences and offers foundations for both unvoiced and voiced pieces. It can be presented by using the following formula for a musical data $X = (x_1, x_2, \dots., x_n)$

$$E = \frac{1}{n}\sum_{i=1}^{n}(x_i)^2 \qquad (3.1)$$

Where E represents the energy, n represents the number of samples in musical data in the database and x_i represents the i^{th}sample in the musical data (X). The x_i is squared in this equation to minimize the summation of the squares of errors that present in each musical sample.

The zero-crossing rate feature is the summation of the crossing per each frame. It is utilized in order to isolate the weak fragments of energy from the quiet background.

This feature can be represented by using the following formula for a musical data $X = (x_1, x_2, \dots., x_n)$

$$ZCR = \frac{1}{n-1}\sum_{i=1}^{n-1} ||\{x_i x_{i+1} > 0\} \qquad (3.2)$$

Where: ZCR represents the zero crossing rate, n is the number of samples in the musical data (X), $||\{A\}$ equals to one when its argument (A) is true, otherwise it equals to zero and $\{x_i x_{i+1} > 0\}$ represents the argument where the multiplication of any time sample and the next time sample must be positive to have $||\{x_i x_{i+1} > 0\}$ equals to 1.

The autocorrelation is the relation between a piece and its own past and future pieces. It represents the similarity among samples as a function of time. This feature is a mathematical tool that is used to find the repeating patterns, as finding the existence of periodic signals that are covered by noises. The correlation between two musical data $X = (x_1, x_2, \dots., x_n)$ and $Y = (y_1, y_2, \dots., y_n)$ can be calculated by using the following formula:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{[\sum(x_i - \bar{x})^2]^{1/2}[\sum(y_i - \bar{y})^2]^{1/2}} \qquad (3.3)$$

Where: x_i represents the i^{th} sample in X, y_i represents the i^{th} sample in Y, $\bar{x}$ and $\bar{y}$ represent the means of both musical data x and y, which can be calculated by using the following two formulas where N represents the number of samples in the musical data, respectively.

$$M = \frac{1}{n}\sum_{i=1}^{n} X_i \qquad (3.4)$$

$$M = \frac{1}{n}\sum_{i=1}^{n} Y_i \qquad (3.5)$$

3.5COMPARISON METHOD

As proposed, the implemented system is divided into two phases; offline and online. In the offline phase, features of the training (known) data are extracted and saved in a database. In the online phase, the features of the testing (unknown) data are compared with those of the training data to find a match between them and then retrieve the most matched data

The features of both the training and testing data are compared based on computing the Minkowski distance between them. The Minkowski distance is the summation of a specific order of the square of the difference between each two corresponding features and the Euclidean distance which is the square root of the square of the difference between each two corresponding features. After that, the sub-melodies which are highly relevant to x_i are recovered from the files database. Minkowski distance is considered as a metric on the Euclidean space which is the generalization of both the Manhattan and Euclidean distances. Minkowski distance is mainly used with p order that has a value 1 or 2. When p equals to 1, it is known as Manhattan distance, while when it equals to 2, it is known as Euclidean distance. The following formula is for the Minkowski distance with p distance order

$$d(X,Y) = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/p} \qquad (3.6)$$

For p equals to 1 or 2, the resultant Manhattan distance and Euclidean distance are:

$$\text{Manhattan distance} = d(X,Y) = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/1} \qquad (3.7)$$

$$\text{Euclidean distance} = d(X,Y) = \left(\sum_{i=1}^{n}(x_i - y_i)^2\right)^{1/2} \qquad (3.8)$$

Where: $X = (x_1, x_2, \ldots \ldots x_n)$, $Y = (y_1, y_2, \ldots \ldots y)$ and p represents the distance order

3.6SYSTEM ACCURACY AND PERFORMANCE

Both the accuracy and performance of the implemented system can be evaluated based on computing the recall measure percentage. The recall measure is the fraction of the related data that are correctly retrieved. Thus, the recall measure depends on recognizing and measuring the relevance of data. The following expression represents the recall measure formula:

$$Recall = \frac{Number\ of\ correctly\ retrieved\ musical\ data}{Total\ number\ of\ data} * 100\% \quad (3.9)$$

The higher the value of the recall measure percentage, the more accurate and precise retrieval system in retrieve the correct data.

CHAPTER FOUR

SYSTEM IMPLEMENTATION AND EVALUATION

This chapter introduces the implementation of retrieval and matching technique for processing the musical data and investigating several music databases by content by using the MATLAB software program. This technique depends on extracting features from musical data and then using them in the Music Information Retrieval (MIR) system. The use of the music features can help in eliminating the restrictions of the traditional techniques which depend on characterizing the music quality properties as well as these features can enhance the performance of music information retrieval algorithms. This technique is applied on a database that contains 200 normal musical data. Each one of data is initially processed based on applying several techniques then divided into two data; training data and testing data.

4.1 MUSIC RETRIEVAL TECHNIQUE BASED ON DWT, LPC AND FILTERING ALGORITHMS

In the processing stage, each one of the used normal musical data is uploaded to the system using the MATLAB where it shows a dialog box that contains a list of all the 200 data and let the user choose the required file. After that, the uploaded file with extension (.wav) is read to return samples from each channel in the file. The returned samples of the file are then plotted in both time domain and frequency domain. The following chart summarizes the proposed stages.

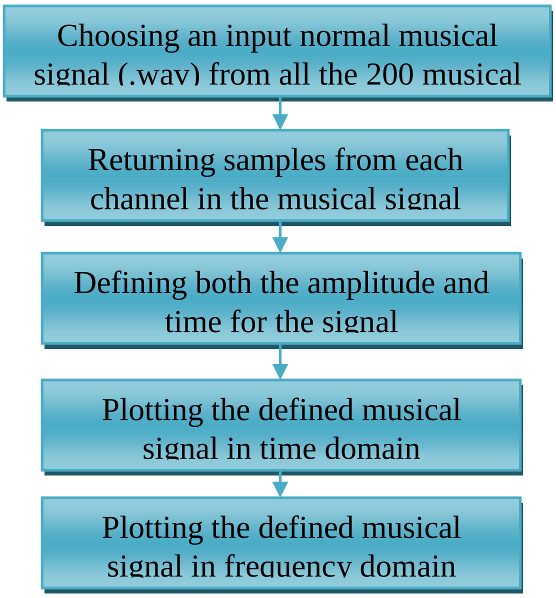

FIGURE 4. 1 STAGES OF PLOTTING A SIGNAL IN BOTH TIME AND FREQUENCY DOMAINS

Figures 4.2 and 4.3 illustrate the time domain and frequency domain of one of the data, respectively.

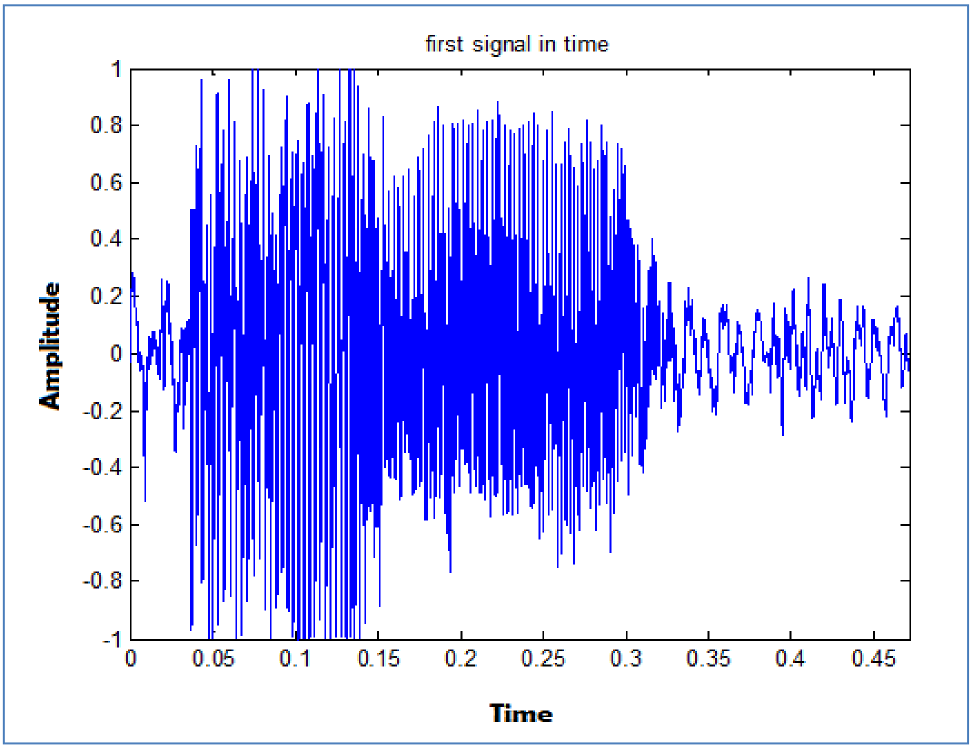

FIGURE 4. 2 FIRST SIGNAL IN TIME DOMAIN OF A VOICE RECORD

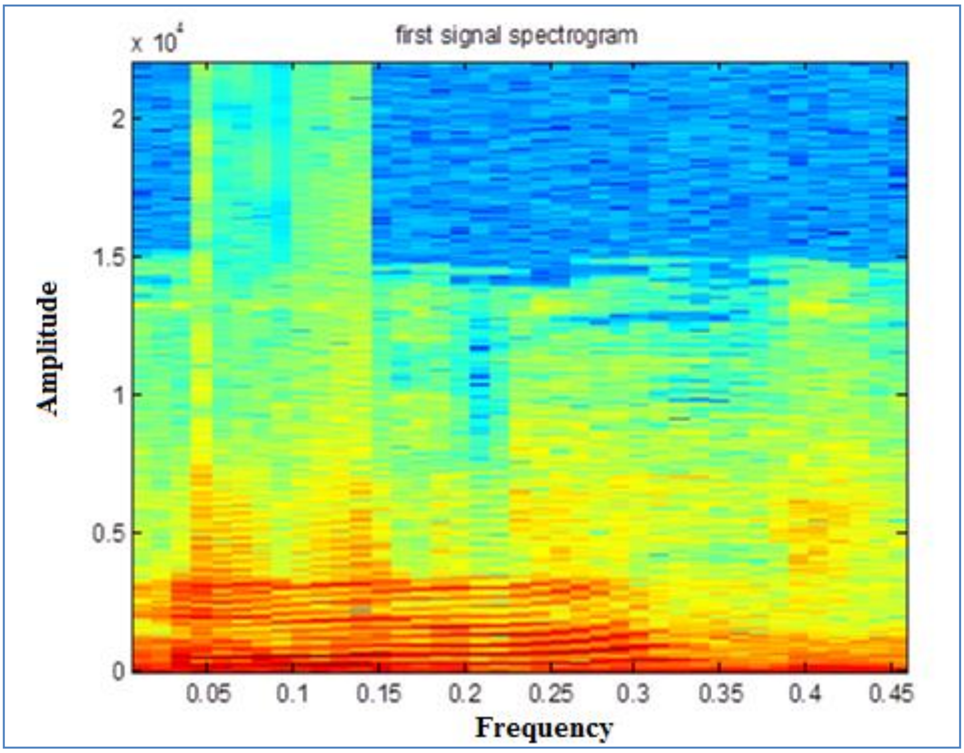

FIGURE 4. 3 FIRST SIGNAL IN FREQUENCY DOMAIN OF A VOICE RECORD

Each one of the data in the read file is played at the default sample rate by using the command sound which converts the vector into sound. This sound is then sent with a specified sample frequency (in hertz) to the speaker that is on the Microsoft Windows.

The values in each sample are supposed to be in the range from -1 to 1. Any value outside this range is deleted. Platforms which support the stereo sounds are used in order to play them when the sample is an nx2 matrix, where the left channel includes values that are in column one, while the right channel includes values that are in column two.

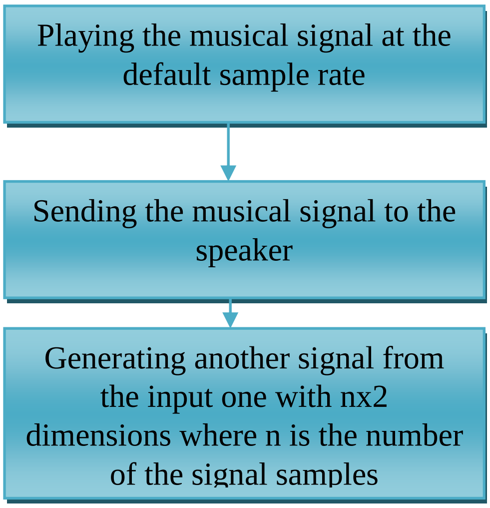

FIGURE 4. 4 STAGES OF GENERATING A NX2 MATRIX MUSICAL SIGNAL

After that, the Discrete Wavelet Transform (DWT) of the input nx2 matrix sound is calculated. The output is an nx2 matrix in which each column in the output is the DWT of the related column in the input. The next process is the filtering which depends on determining a specific frame

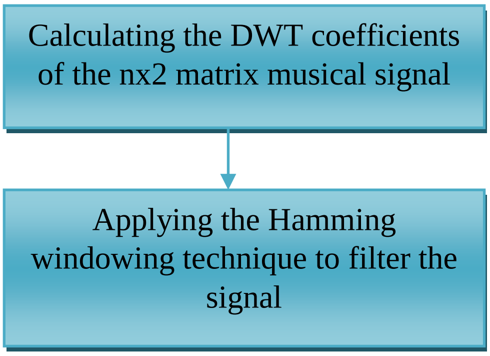

FIGURE 4. 5 DWT CALCULATION AND SIGNAL FILTERATION

The DWT is applied to sample a sound matrix and calculate both the approximation coefficients vector (c_A)and detail coefficients vector(c_D). Thus, for an nx2 sound matrix, the calculated DWT coefficients are:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ . & . \\ . & . \\ a_{n1} & a_{n2} \end{bmatrix} \xrightarrow{DWT} \begin{bmatrix} c_{A,11} & c_{D,12} \\ c_{A,21} & c_{D,22} \\ . & . \\ . & . \\ c_{A,n1} & c_{D,n2} \end{bmatrix}$$

After determining the frame, the Linear Prediction Coefficients (LPC) function is applied in order to find the forward linear analyst coefficients via reducing the calculation error in the least squares sense. This function is used widely in both the speech coding and the filter design. This function finds the coefficients of the specified coefficient value that is in the range from 0 to 1 which determines the recent real valued time series frame depending on the old samples. The specified coefficient value is the prediction filter polynomial order (p). After that, the function filter is used to filter a sequence of data depending on a digital filter that works in both real inputs and complex inputs.

The used filter filters the data that are resulted from the LPC function by using the filter that is demonstrated by the numerator coefficient vector 1. When the value in the vector is not equal to one, then the filter will normalize the coefficients of the filter, while when it is equal to zero, then the filter will give an error. The normalization of the filter coefficients is the process of adjusting the coefficient vector values in which all the vector values become bigger or equal to 0.5 and less than 1. The normalization is done to avoid the overflowing of coefficient and, decrease the filter sensitivity to the quantization and preserve a good filter. This process changes the gain of the filter in order to maintain the magnitude response of the filter. If the filter gain is not changed, then the filter response to a known input will be changed after normalizing the coefficients.

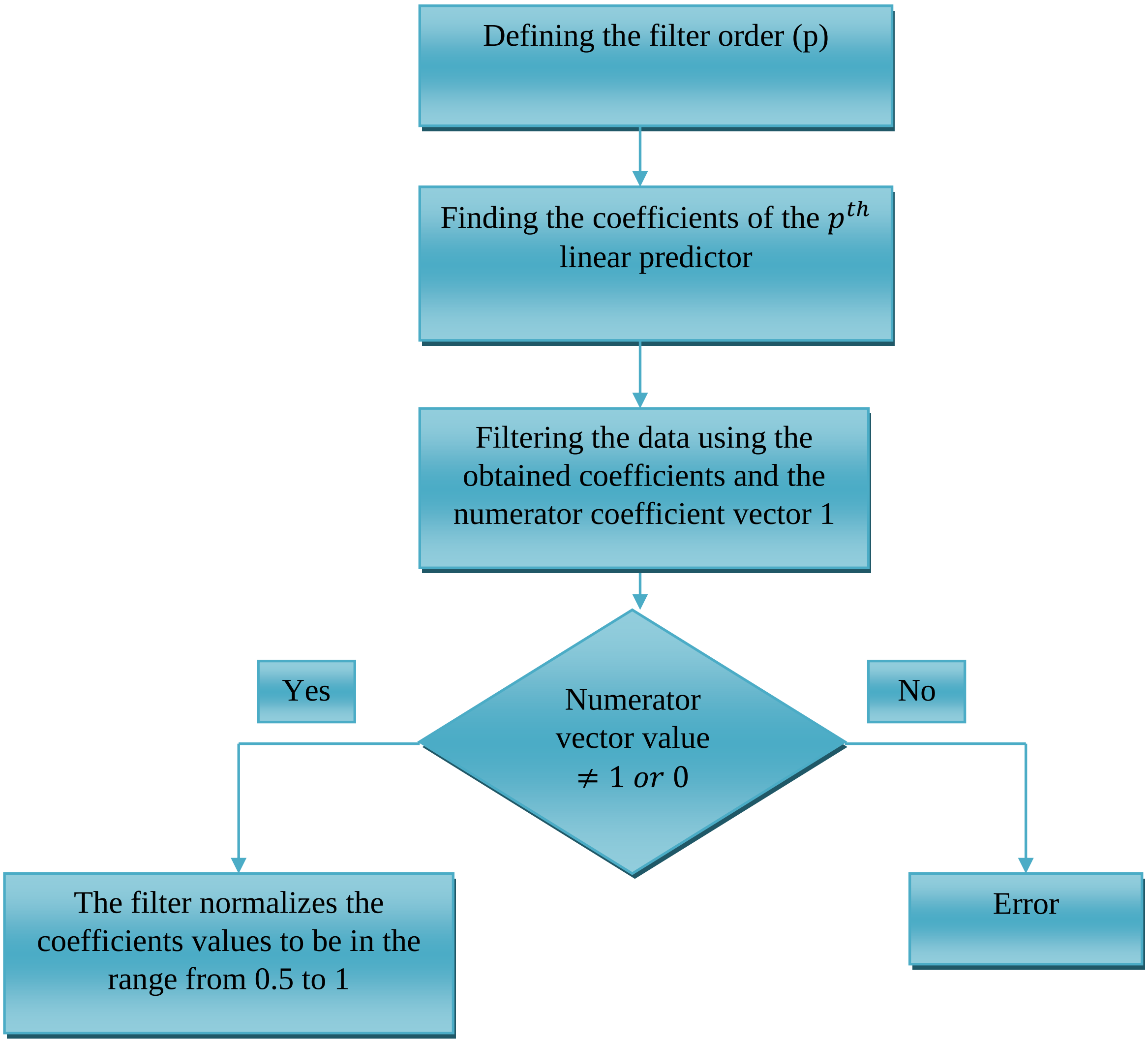

FIGURE 4. 6 STAGES OF APPLYING THE LPC AND FFT TECHNIQUES

After that, both the output of the LPC filtering and the difference between the frame and this output are plotted.

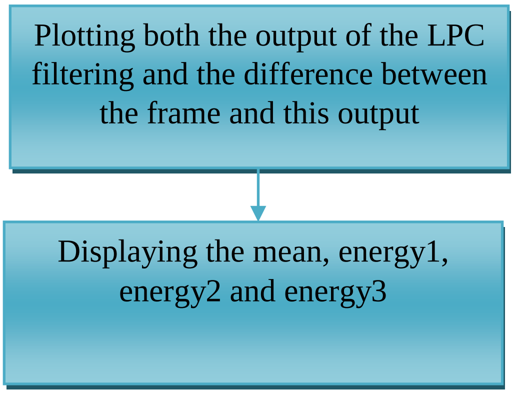

FIGURE 4. 7 STAGES OF PLOTTING THE RESULTANT SIGNAL AND DISPLAYING ITS COEFFICIENTS

The following figure illustrates the output of the LPC filtering and the difference between the frame and this output of one of the 200 data.

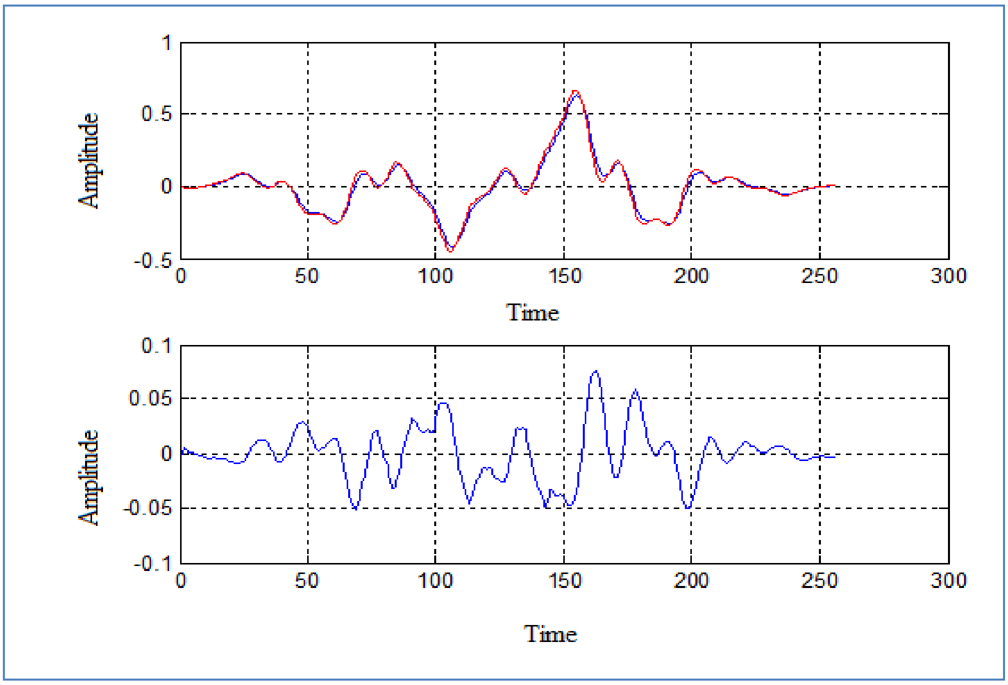

FIGURE 4. 8 THE OUTPUT OF THE LPC FILTERING AND THE DIFFERENCE BETWEEN THE FRAME AND THIS OUTPUT OF A VOICE

The following chart summarizes the proposed music retrieval technique based on DWT for processing the musical data:

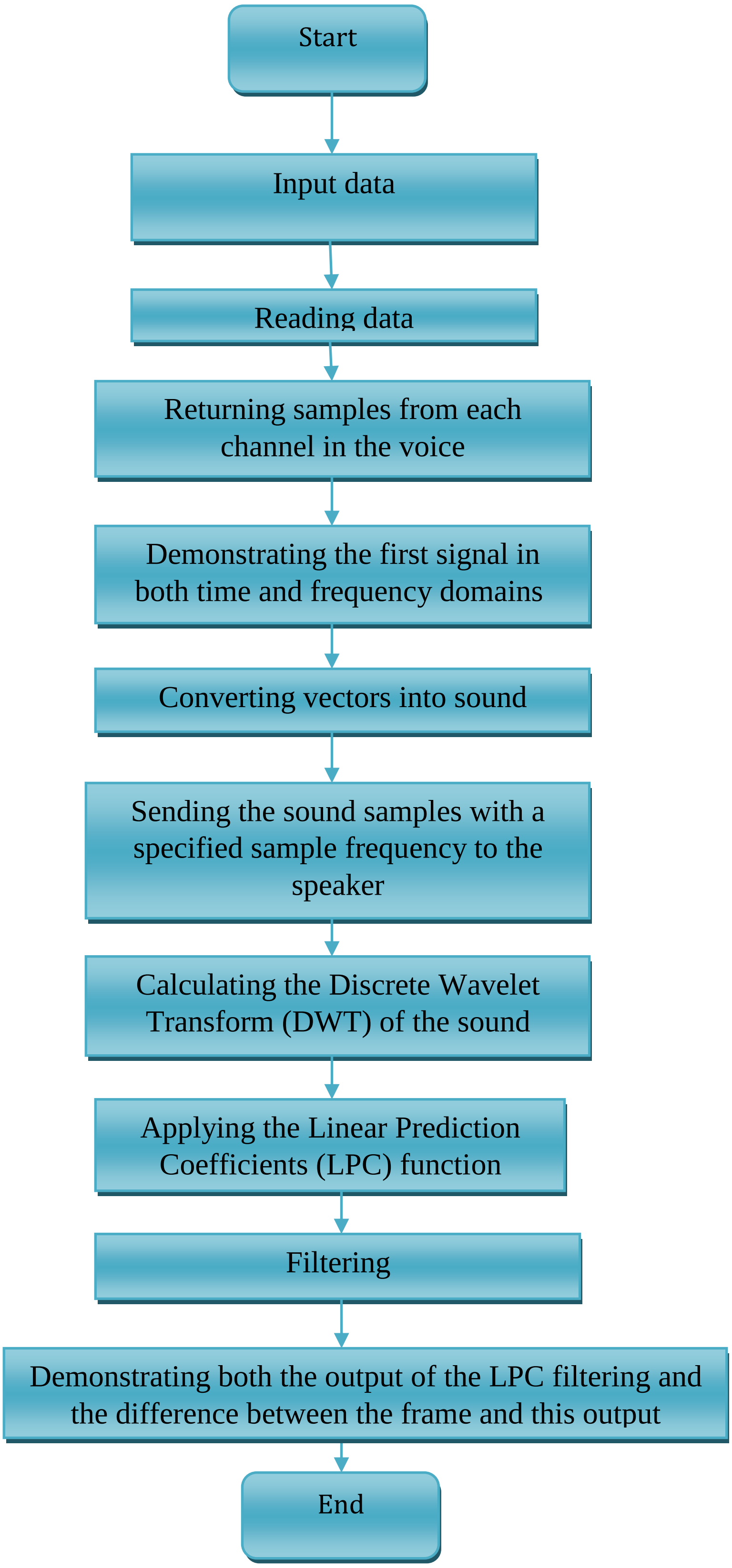

FIGURE 4. 9 PROPOSED MUSIC RETRIEVAL TECHNIQUE BASED ON DWT, LPC AND FILTERING ALGORITHMS

4.2PROPOSED RETRIEVAL SYSTEM AND MATCHING TECHNIQUE

As proposed, 200 data are used to evaluate the proposed retrieval system. Initially, each one of these data is normalized into 20 seconds.

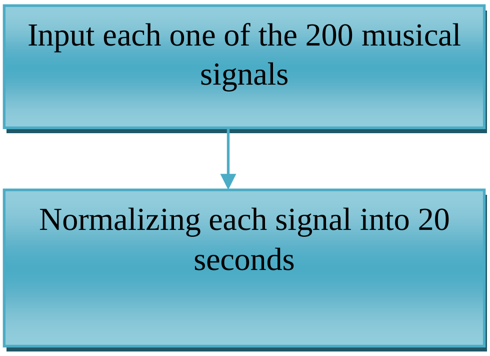

FIGURE 4. 10 NORMALIZATION OF SIGNALS

As example the first data is normalized in time and frequency domains as shown in figures 4.11 and 4.12, respectively.

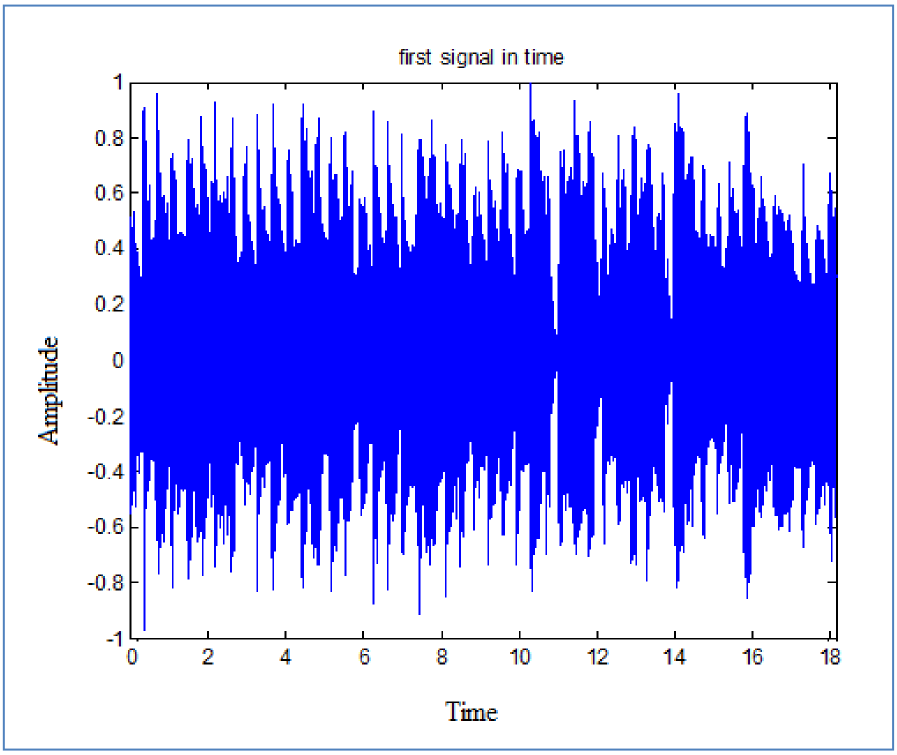

FIGURE 4. 11 NORMALIZATION IN TIME OF A DATA

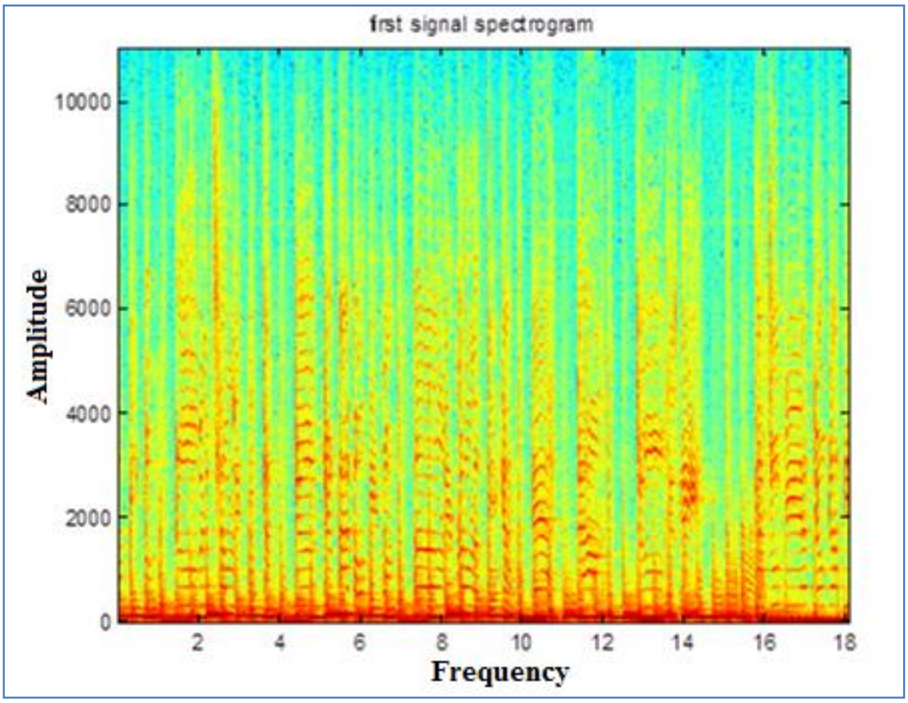

FIGURE 4. 12 NORMALIZATION IN FREQUENCY OF A DATA

After that, each data is divided into two data; training data that represents the first 10 seconds on the data and testing data that represents the other ten seconds. Thus, the proposed signal in figure 4.11 is divided into training and testing data that are demonstrated in figures 4.13 and 4.14 below, respectively.

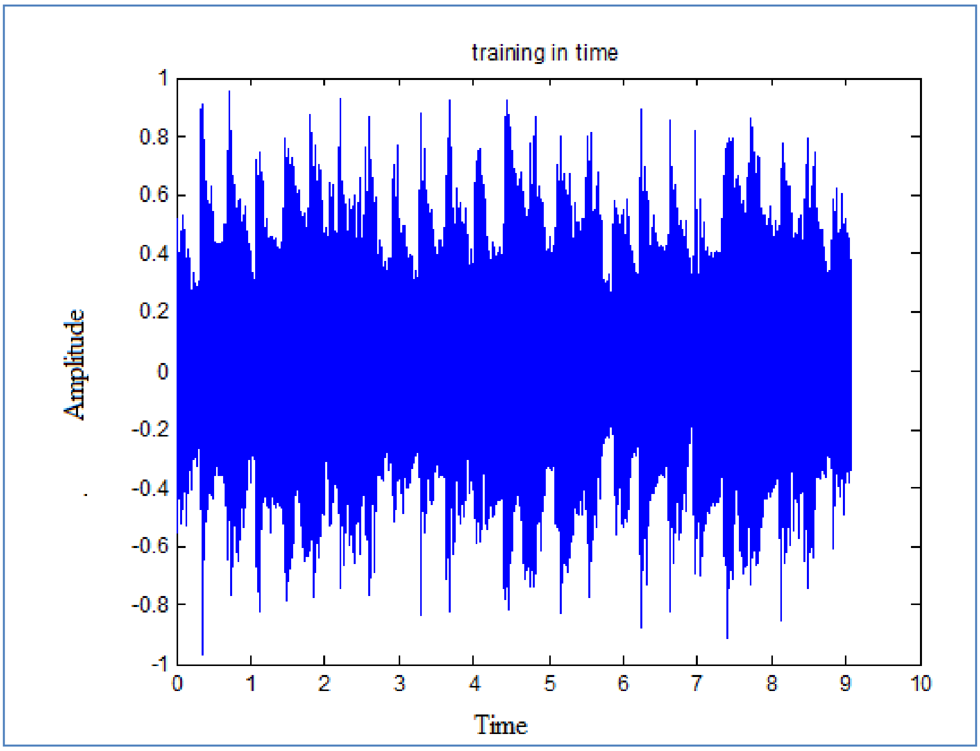

FIGURE 4. 13 TRAINING DATA

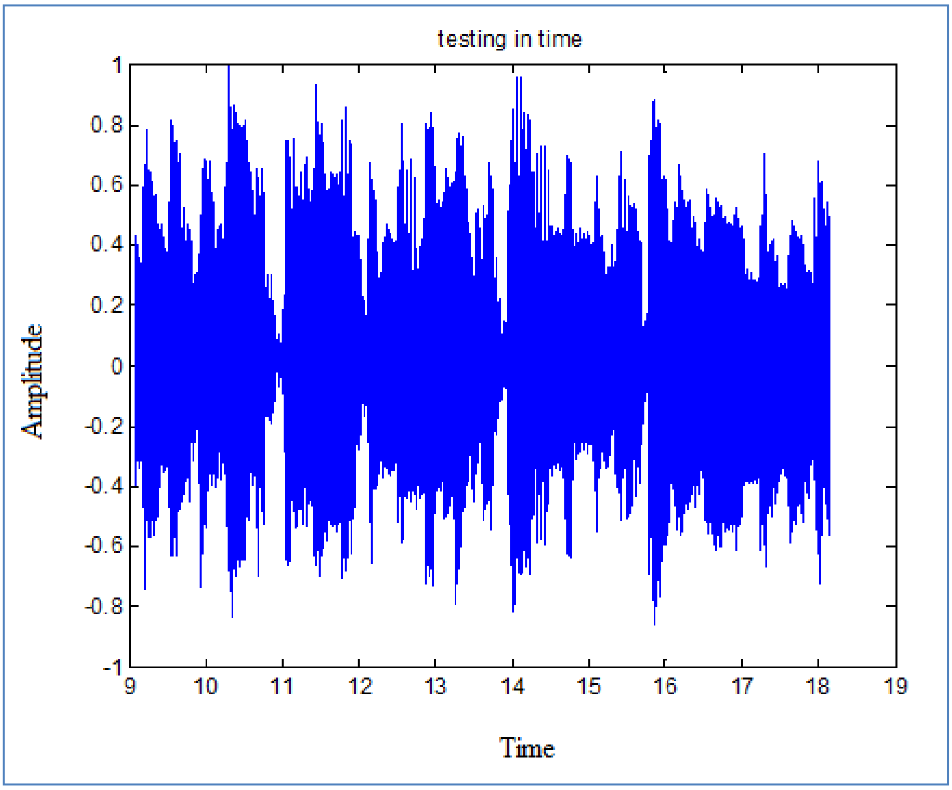

FIGURE 4. 14 TESTING DATA

For the training data, each data is transformed by using the Discrete Wavelet Transform (DWT) into two numerically different bands with the same length; high band and low band. After that, seven features are extracted from each one of the training data, which are: Energy in DWT of the high band, Energy in DWT of the low band, total energy of both bands, zero crossing rate of the high band, zero crossing rate of the low band, autocorrelation of the high band and autocorrelation of the low band. Figure 4.15 illustrates the data that is proposed in figure 4.11 with its resultant energy in DWT of the high (first) band, energy in DWT of the low (second) band and the total energy in DWT.

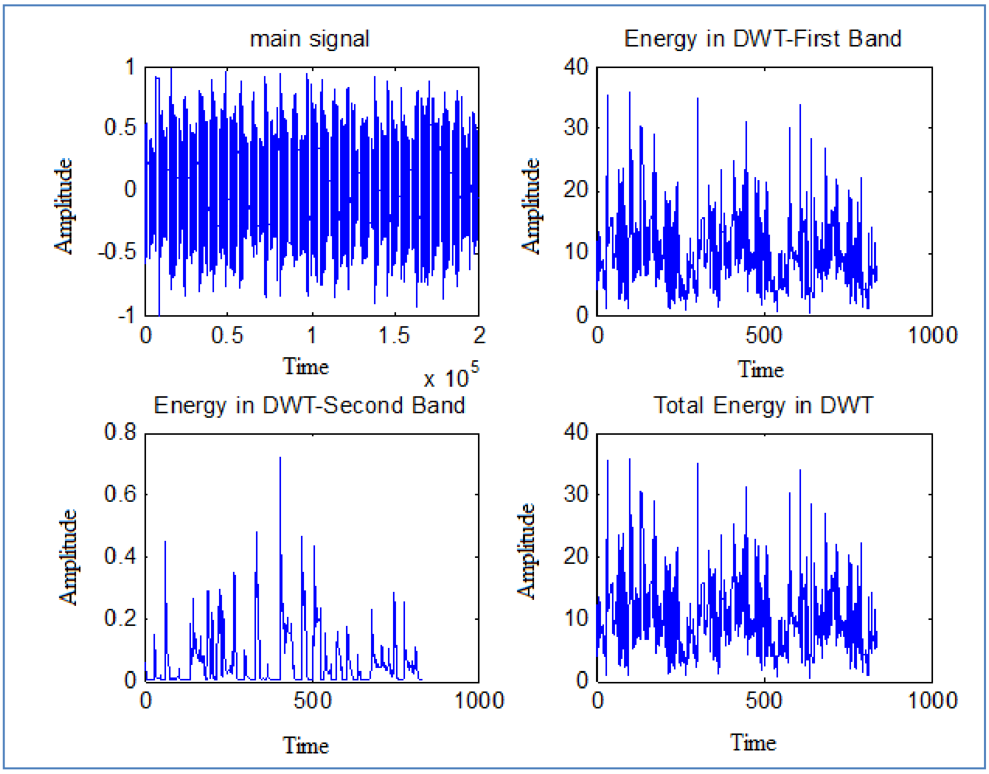

FIGURE 4. 15 MAIN SIGNAL WITH ITS ENERGY IN DWT OF THE FIRST BAND, ENERGY IN DWT OF THE SECOND BAND AND THE TOTAL ENERGY IN DWT

Figure 4.16 demonstrates the data that is proposed in figure 4.5 with its zero crossing rate of the high (first) band, zero crossing rate of the low (second) band, autocorrelation of the high (first) band and autocorrelation of the low (second) band.

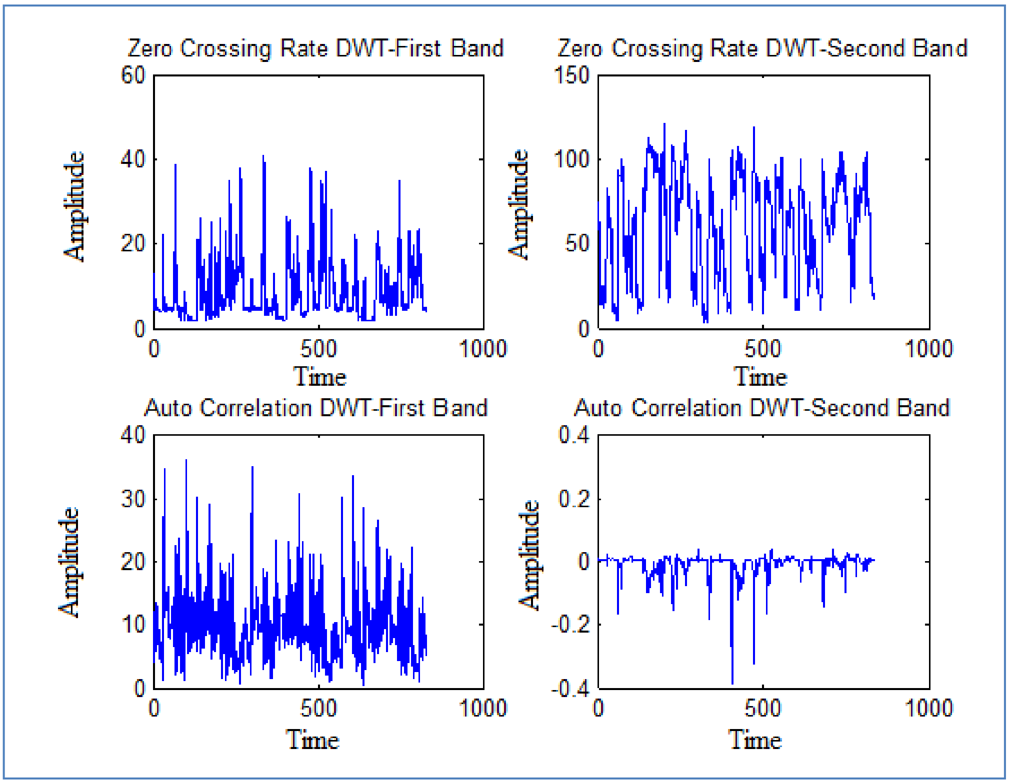

FIGURE 4. 16 MAIN SIGNAL WITH ITS ZERO CROSSING RATE OF THE FIRST BAND, ZERO CROSSING RATE OF THE SECOND BAND, AUTOCORRELATION OF THE FIRST BAND AND AUTOCORRELATION OF THE SECOND BAND

Each one of the proposed seven features of all the data is a vector with dimension 1X833. All the resultant features of the training data are used as a database for the testing data in order to find a match between each test data with its related training data. The table below illustrates a part of the extracted seven features from one of the training data. Each row represents a feature vector.

TABLE 4. 1 A PART OF THE EXTRACTED SEVEN FEATURES FROM TRAINING DATA

F_Features{5,1} <7x833 double>

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	12.1193	8.4857	4.6259	4.0135	7.4724	7.3835	6.6117	13.4450	5.5180	12.2778	8.9050	12.7448	8.1279	8.1240	6.8412
2	0.0621	0.0511	0.0396	0.0306	0.0304	0.0237	0.0143	0.0097	0.0033	0.0054	0.0054	0.0068	0.0041	0.0037	0.0035
3	12.1813	8.5368	4.6655	4.0441	7.5028	7.4072	6.6260	13.4547	5.5213	12.2832	8.9105	12.7515	8.1320	8.1276	6.8447
4	13	9	6	8	4	7	7	4	4	4	5	5	4	4	4
5	73	75	75	70	58	59	64	42	24	13	22	22	25	15	24
6	11.7865	8.1860	4.4411	3.8539	7.3219	7.2517	6.5178	13.3747	5.4927	12.2307	8.8610	12.6892	8.0955	8.0924	6.8117
7	8.4130e-04	0.0068	-0.0017	0.0021	0.0032	0.0038	0.0039	0.0047	0.0024	0.0051	0.0048	0.0062	0.0034	0.0034	0.0031
8															

For the testing data, the same seven features are also extracted for each test data and then compared with the extracted features of all the training data in order to find a match between this test data and its corresponding training data. The following table shows a part of the proposed seven features vectors where each features vector contains 833 elements.

TABLE 4. 2 A PART OF THE EXTRACTED SEVEN FEATURES FROM TESTING DATA

F_TEST <7x833 double>

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	12.1193	8.4857	4.6259	4.0135	7.4724	7.3835	6.6117	13.4450	5.5180	12.2778	8.9050	12.7448	8.1279	8.1240	6.8412
2	0.0621	0.0511	0.0396	0.0306	0.0304	0.0237	0.0143	0.0097	0.0033	0.0054	0.0054	0.0068	0.0041	0.0037	0.0035
3	12.1813	8.5368	4.6655	4.0441	7.5028	7.4072	6.6260	13.4547	5.5213	12.2832	8.9105	12.7515	8.1320	8.1276	6.8447
4	13	9	6	8	4	7	7	4	4	4	5	5	4	4	4
5	73	75	75	70	58	59	64	42	24	13	22	22	25	15	24
6	11.7865	8.1860	4.4411	3.8539	7.3219	7.2517	6.5178	13.3747	5.4927	12.2307	8.8610	12.6892	8.0955	8.0924	6.8117
7	8.4130e-04	0.0068	-0.0017	0.0021	0.0032	0.0038	0.0039	0.0047	0.0024	0.0051	0.0048	0.0062	0.0034	0.0034	0.0031
8															

Figure 4.17 illustrates the test data that is proposed in figure 4.6 with its resultant energy in DWT of the high (first) band, energy in DWT of the low (second) band and the total energy in DWT.

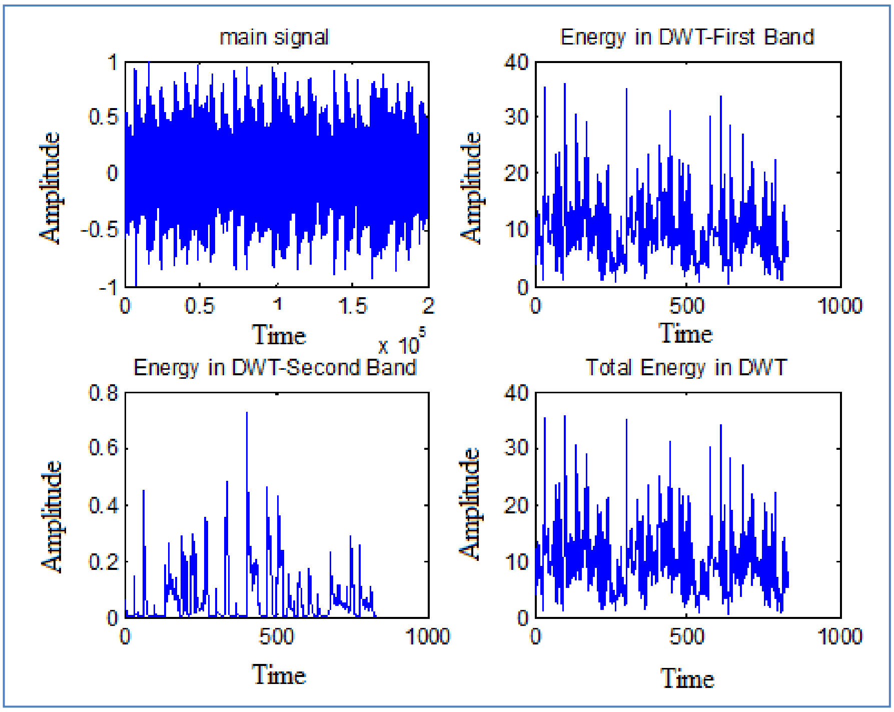

FIGURE 4. 17 MAIN SIGNAL WITH ITS ENERGY IN DWT OF THE FIRST BAND, ENERGY IN DWT OF THE SECOND BAND AND THE TOTAL ENERGY IN DWT

Figure 4.18 demonstrates the data that is proposed in figure 4.6 with its zero crossing rate of the high (first) band, zero crossing rate of the low (second) band, autocorrelation of the high (first) band and autocorrelation of the low (second) band.

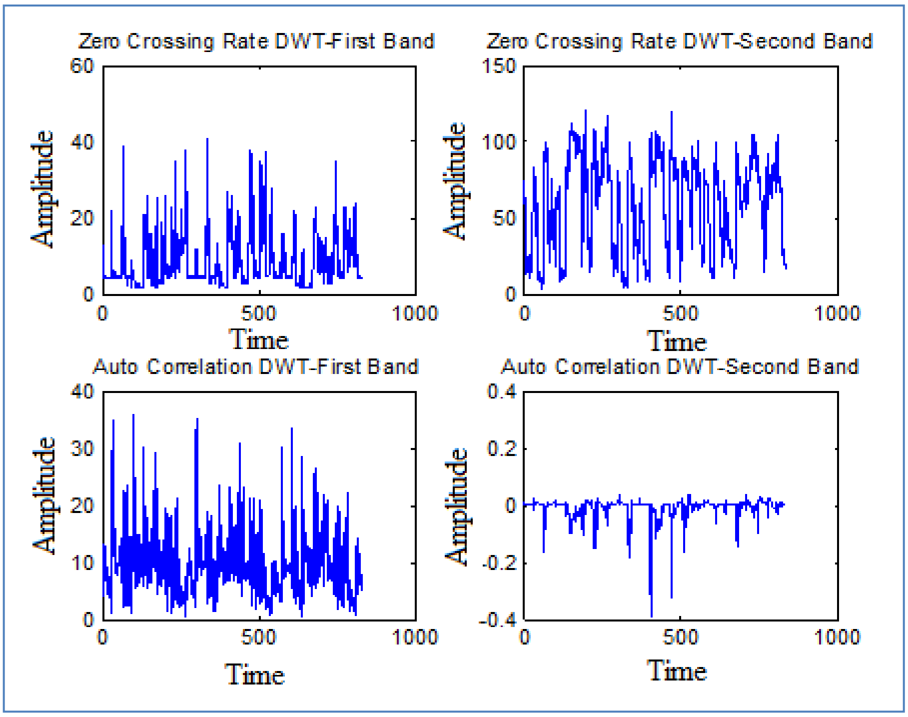

FIGURE 4. 18 MAIN SIGNAL WITH ITS ZERO CROSSING RATE OF THE FIRST BAND, ZERO CROSSING RATE OF THE SECOND BAND, AUTOCORRELATION OF THE FIRST BAND AND AUTOCORRELATION OF THE SECOND BAND

The Minkowski distance is used in order to match the test data with one of the training data based on finding the difference between each feature of the testing data and its corresponding feature of each training data by using the following formula where $\mathrm{X} = (\mathrm{x}_1, \mathrm{x}_2, \dots, \mathrm{x}_\mathrm{n})$ represents the training data with n samples and $\mathrm{Y} = (\mathrm{y}_1, \mathrm{y}, \dots, \mathrm{y}_\mathrm{n})$ represents the testing data with n samples.

$$d(x,y) = \left(\sum_{i=1}^{n} (x_i - y_i)^2\right)^{1/p} \quad (4.1)$$

The used p order in this project is 2. The training data that has the least Minkowski distance with the test data is the matched data that will be retrieved. The results of computing the Minkowski distance between the test data which is a part of the fifth data and all the training data are shown below

TABLE 4. 3 THE RESULTS OF COMPUTING THE MINKOWSKI DISTANCE BETWEEN THE TEST DATA WHICH IS A PART OF THE FIFTH DATA AND ALL THE TRAINING DATA

D_ALL <5x7 double>

	1	2	3	4	5	6	7
1	284.7003	13.0601	282.0777	1.2764e+03	1.2993e+03	300.9593	5.6864
2	266.0126	2.8337	267.3220	279.9554	1.1933e+03	260.3076	1.5102
3	286.6349	3.1007	287.6117	562.1752	908.9213	283.9702	1.7213
4	246.7631	7.0578	245.4955	925.8445	1.2110e+03	256.0701	1.9819
5	0	0	0	0	0	0	0
6							

It can be clearly seen that the computed Minkowski distance between the fifth test data and the training data is zero with the fifth training data which means that there is no difference between the features of both the fifth training and fifth test data. Thus, the system retrieves the fifth training data.

4.3EVALUATION OF THE PROPOSED SYSTEM

In order to evaluate the implemented music retrieval system performance and accuracy, this system is applied on another test data in order to find a match between it and its corresponding training data. The following data which is a part of the third data is used as a test data to evaluate the system.

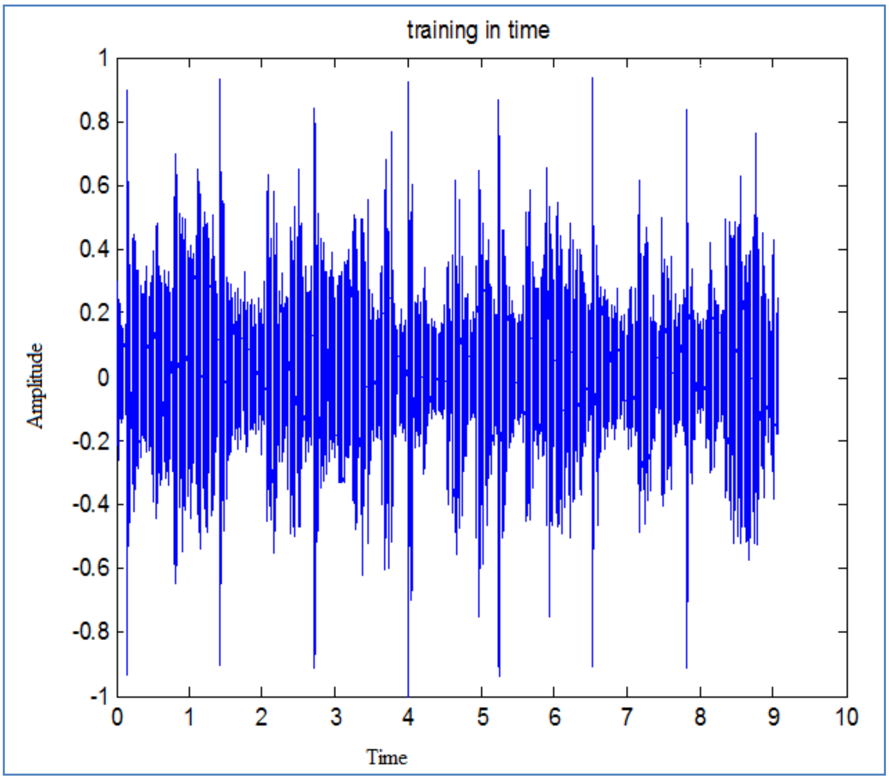

FIGURE 4. 19 TEST DATA

After extracting the features of the test data and all the training data, the Minkowski distance is computed depending on finding the difference between each feature of the testing data and its corresponding feature of each training data. The following table illustrates the computed Minkowski distances between the features of the testing data which is a part of the third data and the training data.

TABLE 4. 4 THE RESULTANT MINKOWSKI DISTANCES

D_ALL <5x7 double>

	1	2	3	4	5	6	7
1	91.0557	13.5515	101.0300	811.1609	1.2009e+03	59.9959	6.1589
2	124.8248	2.0920	126.1567	453.9493	700.8716	115.4106	1.3314
3	0	0	0	0	0	0	0
4	154.3160	7.4533	158.5492	633.4390	984.2764	137.7963	2.3994
5	286.6349	3.1007	287.6117	562.1752	908.9213	283.9702	1.7213
6							

It can be noticed that the calculated Minkowski distances between the features of the third test data and the features of the third training data are zeros. Thus, there is no difference between the features of both data. As a result, the third training data is retrieved. From the evaluation process, one can discover that the implemented music retrieval system can effectively and accurately retrieve the correct data.

In this project, the MATLAB software program with version 7.10.0 (R2010a) is installed in a personal computer with 2.5 GHz and 6 GB to implement and evaluate the proposed music retrieval system. The total consumed time to complete the retrieving process of data is 0.16 seconds. Thus, the proposed system offers enhanced and accurate results in a small period of time.

Both the accuracy and performance of the implemented system are evaluated based on computing the recall measure percentage

$$Recall = \frac{Number\ of\ correctly\ retrieved\ musical\ data}{Total\ number\ of\ data} * 100\%$$

$$= \frac{200}{200} * 100\% = 100\%$$

Thus, the implemented system has the ability to retrieve all the used testing data.

4.4ANALYSIS OF SALT AND PEPPER NOISE EFFECT

This section explores the effect of salt and pepper noise on the training data. This type of noise affect on the data by adding random pixels. The noise variance is an intensity value in the data. This noise can be removed based on using various types of filters. In this section, the effect of the salt and pepper noise on one of the used 200 training data with different variance values is explored. Figure 4.20 below shows the original training data without noise.

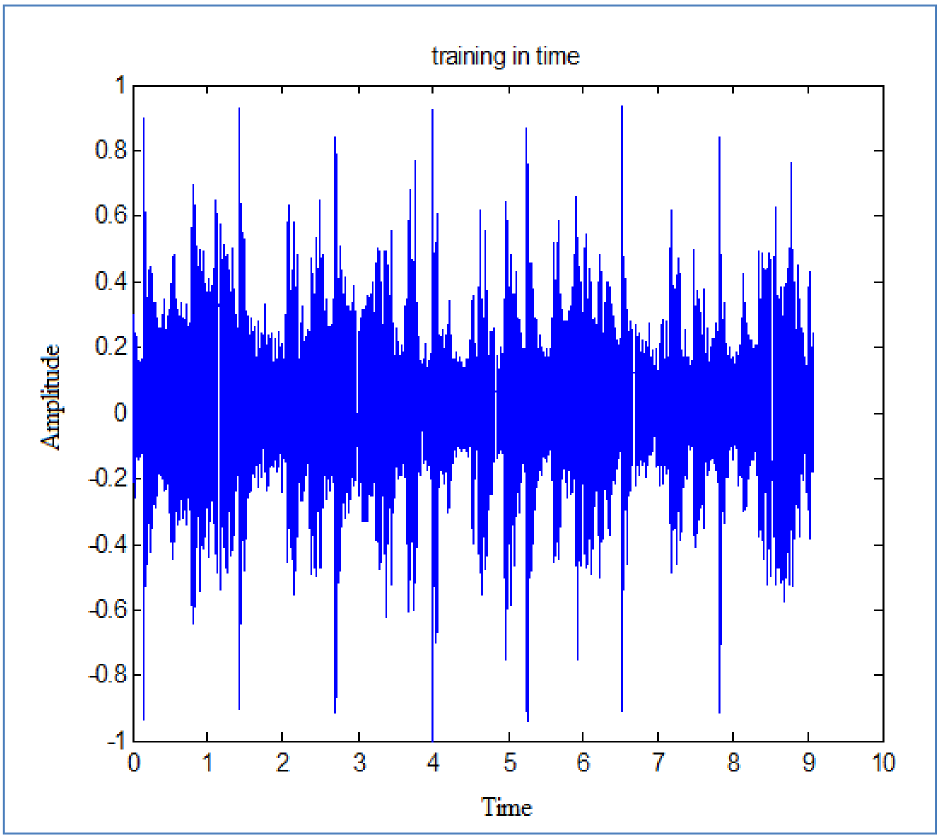

FIGURE 4. 20 ORIGINAL TRAINING DATA WITHOUT NOISE

Table 5 below illustrates both the Signal to Noise Ratios (SNRs) and the values of retrieving percent after applying the salt and pepper noise with various variance values on the proposed data in figure 4.20.

TABLE 4. 5 THE SIGNAL TO NOISE RATIOS (SNRS) AND THE VALUES OF RETRIEVING PERCENT AFTER APPLYING THE SALT AND PEPPER NOISE WITH VARIOUS VARIANCE VALUES

Salt and pepper noise variance value	**retrieving percent**
0.1	85%
0.2	73%
0.3	68%
0.4	58%
0.5	51%
0.6	44%

0.7	31%
0.8	23%
0.9	13%

Figure 4.21 below shows the effect of the proposed noise on the original data with various noise variance values.

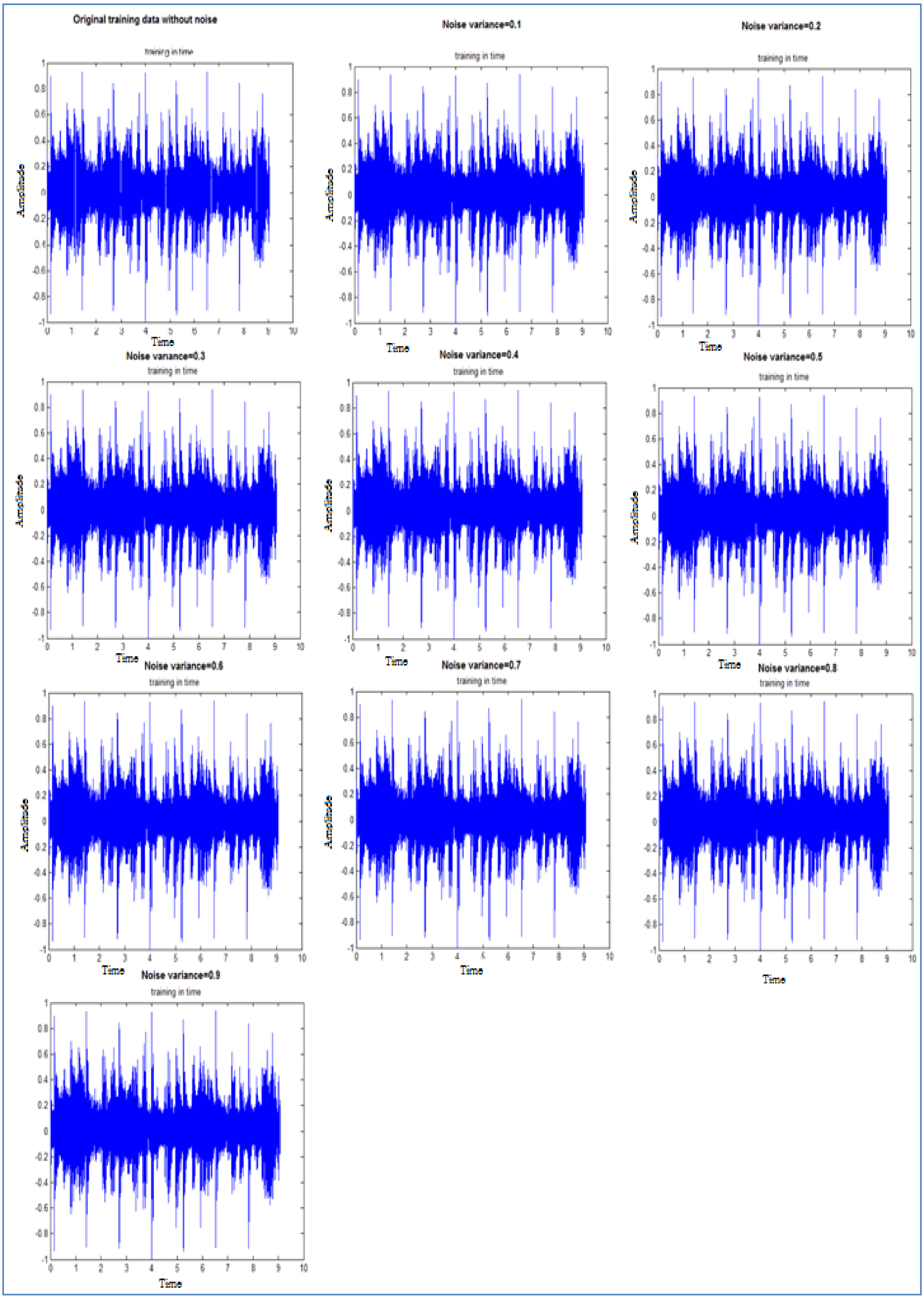

FIGURE 4. 21 EFFECT NOISE ON THE ORIGINAL DATA WITH VARIOUS NOISE VARIANCE VALUES

4.5 COMPARISON WITH PREVIOUS MUSIC RETRIEVAL SYSTEMS

Several researches have been published to improve the efficiency of the music retrieval systems. These researches offer good results in retrieving the musical data but still need more developments.

Rho et al (2008) proposed and implemented an advanced music retrieval system using JAVA. It depends on extracting two main features; hummed query and user query. The hummed query feature represent a set of duration and pitch combinations, while the user query feature is reformulated based on utilizing relevance feedbacks from users. On the other hand, this system depends on applying two advanced algorithms; Windowed Average Energy (WAE) algorithm that enhances the energy dependent methods based on using consistent windows with specified threshold values and the Amplitude based Difference Function (ADF) that enhances the amplitude functions which are used to compute the summation of the signal differences values.

The main stages of the implemented system are: preprocessing based on detecting both the note offset or onset depending on the WAE algorithm and the ADF onset depending on the ADF algorithm, analyzing notes, extracting both the duration and pitch and finally retrieving data. Several experiments were done to assess the implemented system performance. Results proposed that the implemented system offers up to 20–40% retrieval accuracy. On the other hand, the used two algorithms enhance the transcription accuracy to 95%. Although the proposed system offer good performance but it still has a noticed error percentage (RHO. et al, 2008).

Sung et al (2009) proposed another music retrieval system. This system is divided into two main subsystems; client and server. The client subsystem includes two stages; preprocessing stage that reduces the time code gap resulted from the same music contents and features extraction stage that depends on using the Mel Frequency Cepstral Coefficients (MFCC). The server subsystem includes the music server (a file includes the extracted features) and matching stage. This system depends on applying the Dynamic Time Wrapping (DTW) algorithm to find the similarity between data. Results explore that the implemented system can effectively retrieves data with accuracy equals to 97%. On the other hand, this system depends on using features with bad discrimination in similar music sets and cannot detect the complicated starting points.

AI-Qutt et al (2011) proposed an efficient music retrieval system that depends on investigating several statistical and signal processing dependent features. In this work, four statistical features were extracted from musical signals; mean, standard deviation, energy and zero crossing rate, LPC coefficients, pitch, FFT and Haar. The main applied techniques in this system are: the Fast Fourier Transform (FFT), Linear Predicative Coding (LPC), pitch and wavelet analysis. Various melodic similarity measures were used to test features in order to get enhanced search in various musical databases. The obtained results were evaluated using the recall measure. In this work, various types of Daubechies wavelets were used to test the system performance. It was found that the use of the DB2 wavelet has enhanced the recall values, while the use of DB8 wavelet has reduced them. The best obtained recall percentage is 75%.

TABLE 4. 6 COMPARISON BETWEEN SYSTEMS

Music retrieval system	**Extracted features and used database**	**Used algorithms**	**Accuracy**	**Dependency on normalization process**	**Complexity**	**Results**
Rho et al (2008) system	hummed query and user query It uses large database	Windowed Average Energy (WAE) algorithm and Amplitude based Difference Function (ADF) algorithm	High	Independent	Slightly complex due to the complicated algorithms that used in this system	The implemented system offers up to 20–40% retrieval accuracy. The used two algorithms enhance the transcription accuracy to **95%.**
Sung et	Mel	Dynamic	High	Independent	Slightly	the implemented

al (2009) system	Frequency Cepstral Coefficients (MFCC) It uses large database	Time Wrapping (DTW) algorithm			complex due to the complicated algorithm that used in this system	system can effectively retrieves data with accuracy equals to **97%**
AI-Qutt et al (2011) system	Statistical features; Mean, standard deviation, energy and zero crossing rate, LPC coefficients, pitch, FFT and Haar It uses large database	Fast Fourier Transform (FFT), Linear Predicative Coding (LPC), pitch and wavelet analysis	Low	Independent	Simple	The implemented music retrieval system can effectively retrieve music with recall percentage equals to **75%**
Proposed system	energy in DWT of the high band, Energy in DWT of the low band, total energy of both bands, zero crossing rate of the high band, zero crossing rate	Discrete Wavelet Transform (DWT), Linear Predictive Coding (LPC) and filtering function	Highest accuracy	Dependent	Simple	The implemented music retrieval system can effectively and accurately retrieve the correct data for **100%** the used testing data.

	of the low band, autocorrelation of the high band and autocorrelation of the low band It uses large database					

Thus, the proposed system comes to solve several problems of the traditional system, enhance the music retrieval accuracy and improve the system performance.

CHAPTER FIVE

CONCLUSION AND FUTURE WORK

5.1CONCLUSION

This project proposed the implementation of a robust retrieval and matching technique for retrieving several musical data using the MATLAB software program. The implemented system depends on extracting some music features from the data which in turn solve the traditional techniques problems that based on characterizing the music quality properties.

The implemented system is applied on a database that contains 200 different musical data. These data are normalized into 20 seconds. The main stages of the implemented system are: dividing each one of the used 200 data into training and testing data where the duration of each one is 20 seconds, pre-processing the resultant data based on applying three algorithms; Discrete Wavelet Transform (DWT), Linear Predictive Coding (LPC) and filtering function, extracting seven features of both data; energy in DWT of the high band, Energy in DWT of the low band, total energy of both bands, zero crossing rate of the high band, zero crossing rate of the low band, autocorrelation of the high band and autocorrelation of the low band, comparing the extracted features of both the training and testing data based on computing the Minkowski distance and finally retrieving the most matched data.

The training data that has the least Minkowski distance with the test data is the matched data that will be retrieved. Results explore that the Minkowski distance between the matched data equals to zero. It was found that the implemented music retrieval system can effectively and accurately retrieve the correct data based on the chosen testing data.

5.2FUTURE WORKS

This project introduces the implementation of advanced music retrieval system. This system offers high accuracy and enhanced performance in retrieving the matched musical data. On the other hand, the implemented system can be developed in the future based on applying it in a large database, extracting more features and recognizing instruments that are playing in a mixture sound without dividing them.

REFERENCES

- AIQUTT, M.M, HAMAD, A.M, SALEM, M.A and ABDEL-AZIZ, M.H. 2011. Performance Comparison for Feature Selection in Musical Information Retrieval. pp. 231-236
- AUCOUTURIER, J and PACHET, F. 2002. Music similarity measures: What's the use? In *Proc. Int. Conf. Music Information Retrieval*.
- AUCOUTURIER, J. PACHET, F. and SANDLER, M. 2005. The Way It Sounds: Timbre Models for Analysis and Retrieval of Music Signals, IEEE transactions on multimedia, **7**(6).
- BIRMINGHAM, W, DANNENBERG, R AND PARDO, B. 2006. Query by humming with the vocalsearch system. Communications of the ACM 49 (8). pp. 49–52.
- CANO, P, BATLLE, E, KALKER, T, and HAITSMA, J. 2005. A review of audio fingerprinting. The Journal of VLSI Signal Processing. 41 (3). pp. 271–284.
- CASEY, M. A, VELTKAMP, R, GOTO, M, LEMAN, M, RHODES, C and SLANEY, M. 2008. Content Based Music Information Retrieval: Current Directions and Future Challenges. *IEEE*
- CASTAN, C. 2006. Music Notation Formats. [Online]: available: http://www.music-notation.info
- DAGUANNO, A, HAUS, G, PINTO, A and VERCELLESI, G. 2007. Tools for music information retrieval and playing.
- EBERT, D. S. ROHRER, R. M. SHAW, C. D. PANDA, P. KUKLA, J. M and ROBERTS, D. A. 1999. Procedural shape generation for multi-dimensional data Visualization. In: Data Visualization '99. pp. 3–12.
- FOOTE, J. T. 2008. Content-based retrieval of music and audio, in *Proc. SPIE Multimedia Storage and Archiving Systems II*.[109] K. Seyerlehner, G. Widmer, and P. Knees, "Frame-level audio similarity—A codebook approach," in *Proc. Int. Conf. Digital Audio Effects*. pp. 138–147.
- FU, Z. LU, G.TING, K. M and ZHANG, D. 2011. A Survey of Audio-Based Music.
- FU, Z. LU, G. TING, K. M and ZHANG, D. 2011. Music classification via the bag-of-features approach.

- GOLDSTEIN, T and OSHER, S. 2008. The split Bregman method for L1 regularized problems.
- HOFFMAN, M. BLEI, D and Cook, P. 2008.Content-based musical similarity computation using the hierarchical dirichlet process. in *Proc. Int. Conf. Music Information Retrieval*.
- HURON, D. 2000. *Proceedings of the 1 st Society for Music Information Retrieval Conference (lSMIR).*
- KAMINSKAS, M and RICCI, F. 2012. Contextual music information retrieval and recommendation: State of the art and challenges. *Elsevier: Computer science review*.
- KOLAKOWSKA, A. 2008. *1st International Conference on Information Technology*, Gdansk, Poland
- KOSUGI, N. 2004. A Study on Content-based Music Retrieval with Humming
- LAMERE, P. 2008. SOCIAL TAGGING AND MUSIC INFORMATION RETRIEVAL. *Sun Microsystems Inc*
- LEW, M. S. SEBE, N. DJERABA, C and JAIN, R. 2006. Content-based multimedia information retrieval: state of the art and challenges. ACM Transactions on Multimedia Computing, Communications, and Applications. **2**(1). pp. 1–19.
- LIDY, T and RAUBER, A. 2005. Evaluation of feature extractors and psychoacoustic transformations for music genre calassification. in Proc. Int. Conf. Music Information Retrieval.
- LINDENBAUM, O. MASKIT, S. KUTIEL, O. and NAVE, G. 2010. *Musical Features Extraction for Audio-based Search*, IEEE 26-th Convention of Electrical and Electronics Engineers in Israel.
- LOGAN, B and SALOMAN, A. 2001. A music similarity function based on signal analysis, in *Proc. Int. Conf. Multimedia Expo*.
- LOGAN, B. 2000. Mel frequency cepstral coefficients for music modeling. In: ISMIR ’00: international symposium on music information retrieval.
- LOMONT, C. 2010. The Fast Fourier Transform
- LU, Y. 2008. STATISTICAL LEARNING FOR MULTIMEDIA INFORMATION RETRIEVAL AND THEIR APPLICATIONS. Submitter to the university of Texas at San Antonio

- MANDEL, M and ELLIS, D. 2005. Song-level features and SVMs for music classification. In Proc. Int. Conf. Music Information Retrieval.
- MAYBURY, M. T.1997. Editor, Intelligent Multimedia Information Retrieval.
- MIOTTO, R and ORIO, N. 2010. *11th International Society for Music Information Retrieval Conference*.
- MOSA, G. S and ALI, A. A. 2009. Arabic Phoneme Recognition using Hierarchical Neural Fuzzy Petri Net and LPC Feature Extraction. *Signal Processing: An International Journal (SPIJ).* **3** (5). pp.161-171
- ORIO, N. 2006. Music Retrieval: A Tutorial and Review
- PACHET, F and CAZALY, D. 2000. A taxonomy of musical genres, in: Proc. Content-Based Multimedia Information Access, RIAO. pp. 1238–1245.
- PAMPALK, E. DIXON, S and WIDMER, G. 2003. On the evaluation of perceptual similarity measures for music, in *Proc. Int. Conf. Music Information Retrieval*
- PICKNES, J, BELLO, J.P, MONTI, G, SANDLER, M, CRAWFORD, T, DOVEY, M, and BYRD, D. 2003. Polyphonic Score Retrieval Using Polyphonic Audio Queries: A Harmonic Modeling Approach. Journal of New Music Research. 32. pp.223–236CASEY, M. 2006. Audio Tools for Music Discovery. [online]: available: http://www.methodsnetwork.ac.uk /redist/pdf/casey.pdf
- POPE, S. T. HOLM, F. and KOUZNETSOV, A. 2004. *Feature Extraction and Database Design for Music Software*, Proc. 2004 Int'l Computer Music Conference, U. of Miami, 11/04.
- REN, F and BRACEWELL, D. B. 2009. Advanced Information Retrieval.
- SCHIIUBLE, P. 1997. Multimedia Information Retrieval: Content-Based Information Retrieval from Large Text and Audio Databases
- SEYERLEHNER, K. WIDMER, G and KNEES, P. 2008. Frame-level audio similarity—A codebook approach," in *Proc. Int. Conf. Digital Audio Effects*.
- STEPHEN, D.J. 2003. Music information retrieval. *Annual Review of Information Science and Technology*. **37**. pp.295-340.
- SUNG, B. KOO, K. KIM, J. JUNG, M. KWON, J and KO, I. 2009. *Effective Digital Music Retrieval System through Content-based Features*, World Academy of Science, Engineering and Technology 50 2009.

- TSAI, W.H and WANG, H.M. 2006. Automatic singer recognition of popular music recordings via estimation and modeling of solo vocal signals. *Trans. Audio, Speech, and Language Processing*. **14** (1). pp. 330–341
- TZANETAKIS, G and COOK, P. 2002. Musical genre classification of audio signals, IEEE Trans. Speech Audio Process. **10**(5). pp. 293–302.
- ULAS¸ BAˇGCI and Engin Erzin. 2007. Automatic Classification of Musical Genres Using Inter-Genre Similarity, **14**(8).
- WANG, J.Z, BOUJEMAA, N, BIMBO, A.D, GEMAN, D, HAUPTMANN, A.G and TESIC, J. 2006. *Proceedings of the 8th ACM international workshop on Multimedia information retrieval*. Santa Barbara, California, USA.
- WIERING, F. 2006. Can Humans Benefit from Music Information Retrieval?
- THIANG and WIJOYO, S. 2011. Speech Recognition Using Linear Predictive Coding and Artificial Neural Network for Controlling Movement of Mobile Robot. *International Conference on Information and Electronics Engineering*. pp. 179-183
- RHO, S, HAN, B. J, HWANG, E and KIM, M. 2008. MUSEMBLE: A novel music retrieval system with automatic voice query transcription and reformulation. *The Journal of Systems and Software* **81**. pp.1065–1080
- SUNG, B, KOO, K, KIM, J, JUNG, M. B, KWON, J and KO, I. 2009. Effective Digital Music Retrieval System through Content-based Features. *World Academy of Science, Engineering and Technology*. **26**. pp. 721-762