J Inf Process Syst, Vol.14, No.1, pp.32~55, February 2018
https://doi.org/10.3745/JIPS.04.0054

ISSN 1976-913X (Print)
ISSN 2092-805X (Electronic)

JOURNAL OF INFORMATION PROCESSING SYSTEMS JIPS

A Comparative Analysis of Music Similarity Measures in Music Information Retrieval Systems

Kuldeep Gurjar* and Yang-Sae Moon*

Abstract
The digitization of music has seen a considerable increase in audience size from a few localized listeners to a wider range of global listeners. At the same time, the digitization brings the challenge of smoothly retrieving music from large databases. To deal with this challenge, many systems which support the smooth retrieval of musical data have been developed. At the computational level, a query music piece is compared with the rest of the music pieces in the database. These systems, music information retrieval (MIR systems), work for various applications such as general music retrieval, plagiarism detection, music recommendation, and musicology. This paper mainly addresses two parts of the MIR research area. First, it presents a general overview of MIR, which will examine the history of MIR, the functionality of MIR, application areas of MIR, and the components of MIR. Second, we will investigate music similarity measurement methods, where we provide a comparative analysis of state of the art methods. The scope of this paper focuses on comparative analysis of the accuracy and efficiency of a few key MIR systems. These analyses help in understanding the current and future challenges associated with the field of MIR systems and music similarity measures.

Keywords
Content-Based Music Retrieval, MIR System, Music Information Retrieval Survey, Music Similarity Measures

1. Introduction

The preliminary ambition in the field of arts has always been to be appreciated, and hence the greatest challenge is to reach the maximum number of audience. To deal with this challenge, information technology (IT) provides different solutions for different arts. Digitization being one of the proposed solutions do not suit the arts like painting and sculpture but works almost perfectly for music. It has improved the music industry regarding storage, accessibility, and most importantly business. For example, iTunes is the world's largest music vendor that offers 43 million songs.

Music information retrieval (MIR) is primarily concerned with the three following steps. The first is the extraction and inference of meaningful and computable features from the audio signal, symbolic representation or external sources such as web pages. The second is the indexing of music using these features.

The third is the development of different search and retrieval schemes such as content-based search, music recommendation systems, or user interfaces for browsing large music collections, as defined by

Manuscript received November 13, 2017; accepted December 18, 2017.
Corresponding Author: Yang-Sae Moon (ysmoon@kangwon.ac.kr)
* Dept. of Computer Science, Kangwon National University, Chuncheon, Korea (dkekuldeep@gmail.com, ysmoon@kangwon.ac.kr)

Downie [1]. Owing to a rapid development of application areas, the research in the field of MIR systems has experienced a constant upward trend.

The MIR research is an emerging yet established field with the application areas such as music browsing, plagiarism detection, music recommendation systems, and musicology [2]. However, if we see the importance of music in our daily lives, the research field of MIR systems seems a bit delayed, as it started less than twenty years ago. Based on this investigation, in this paper we present a brief history and evaluation of MIR first to elaborate further on it.

The scope of the paper focuses on MIR systems and its development, keeping aside the musical technicalities. In MIR systems, our primary focus is on the music similarity comparison with a brief introduction about extracting features from audio. The rest of the paper is organized as follows. Section 2 introduces the history and evolution of MIR systems. Section 3 describes the state of the art methods of MIR systems. Section 4 presents components of an MIR system. Section 5 proposes application areas of MIR systems; Section 6 provides a comparative analysis of the methods. Finally, we summarize the paper in Section 7 with future challenges of the field.

2. History and Evolution of MIR Systems

In this section, we describe the history, evolution, and various challenges been faced by the MIR research area over time. We also present how increasing computational power is boosting the field of signal processing techniques and its related applications. Early MIR research more focused on working with symbolic representations of music pieces (i.e. a structured, digital representation of musical scores such as music instrument digital interface [MIDI]). During the early 2000s, due to advances in computation technology, a new era of signal processing techniques began which aimed directly at music audio signals. It allowed processing of not only music scores (mainly available for Western Classical music) but also of different types of recorded music. That was achieved by deriving different music qualities (e.g. rhythm, timbre, melody, or harmony) from the audio signal itself which is still claimed to be a frequently pursued endeavor in today's MIR research as stated by Casey et al. [2].

In the mid to late 1960s, some very significant work of outlining the criteria for an MIR system was proposed by Kassler [3,4] and Lincoln [5]. These research works made it possible to index musical themes using computers and gave direction to further researches. Later research works proposed by Deutsch [6] and Byrd [7] addressed tune recognition and improvement of music notation software, respectively. Also, Mongeau and Sankoff [8] proposed their edit-distance dynamically programmed approach, which works as the base for the modern theory of MIR systems. Until the mid-1990s, various research works were proposed to further refine the idea of eliminating the human component for maintaining consistency of audio transcription and searching; a bulk of this work targeted the speech transcription domain, with some potential findings applicable to topics in MIR [9]. Over the last 15 years, however, there has been a tremendous increase in the number of research works on the subject, and that can be accounted for the following factors:

- Emergence of better MIDI standards: MIDI is a protocol designed to record and play music digitally. As a mathematical based musical file format, if has simplified musical analysis and categorization for database storage. MIDI has led many to reconsider the use of the MIDI file format as a basis for musical information sharing and analysis.

- Increased accessibility: There has been an increase of freely and publicly available repositories of downloadable music via the Internet. And this has increased the potential utility of musical databases by making the information available for searching.
- Development of compression formats: Compression formats such as Ogg Vorbis, and more significantly MP3, have reduced musical file sizes making it possible to reasonably store a large number of songs within a single database.
- *Increased affordability*: With the all-round development in the field, it is becoming more and more affordable as the per-byte cost of data storage media continues to improve.
- *Increased interest in MIR*: With the emergence of new and exciting application areas, MIR got the attention of a many researchers. This was partially sparked by a widely-cited and ground-breaking papers such as [10-12]. These works outlined various frameworks for an online music information retrieval system with appropriate query mechanisms. Moreover, they also addressed the future challenges of MIR systems. The insight provided by these papers led many to invest in MIR research areas.

According to Freed [13], the concept of retrieving music through metadata introduced in the late 1990s is still highly successful in the field. However, there are two basic problems with textual metadata. The first problem is the excessiveness of data, which further leads to the problem of increased cost to retrieve and maintain that data. The second problem is the requirement of maintaining consistent or uniformed expressions. As different people and with their various perceptions create a variety of descriptions which lead to variation in concept encodings, and this adversely impacts the search performance. Furthermore, not only the notions of music description but also the concept of music similarity are vague and hence human agreement on the similarity between two music pieces is bounded at about 80% as stated in several past works [13-17]. The descriptions represent opinions, so editorial supervision of metadata is paramount [12].

In addition to the metadata challenge mentioned above, MIR had some additional challenges to be discussed below. In general, the music information consists of pitch, temporal, harmonic, timbre, editorial, textual, and bibliographic facets. Accessing the multifaceted information is not the only challenge faced by MIR research. Developers and evaluators must also constantly keep track of the many ways in which music is represented to overcome the "multi-representational challenge". Due to time immemorial, various cultures and subcultures have created their unique style of expression via music, and this cultural expression leads to a "multi-cultural challenge". In addition, every individual interacts with music and MIR systems in a different manner. And the process of comprehending and analyzing these interactions forms "multi-experiential systems". We discuss the challenges of the MIR field in Section 7. For further details regarding the challenges faced in MIR research, readers are referred to Byrd and Crawford [10] and Downie [1].

Recently, we have witnessed a meaningful shift from system-centric designs to user-centric designs of MIR, in both models and evaluation procedures as mentioned by Casey et al. [2]. User-centric strategies account for different factors in the perception of music qualities, especially music similarity. Application areas such as music recommendation and playlist generation aim at providing a customized listening experience for a user. In the next section, we present the functionality of MIR systems, focusing mainly on the similarity comparison methods.

3. MIR: Functionality and State-of-the-Art Methods

In this section, we explain the development and functionality of MIR systems followed by the state-of-the-art methods. The definition of MIR systems has refined gradually by time, here are a few important definitions proposed by various researchers. Downie [18] defines MIR as a multidisciplinary research endeavor that strives to develop innovative content-based searching schemes, novel interfaces, and evolving networked delivery mechanisms to match the world's vast store of music accessible to all. According to Demopoulos and Katchabaw [19], the ultimate goal of MIR is to identify various words and properties contained within musical selections, storing these data within a musical database, and providing a query mechanism to specify searches over the database to retrieve the musical documents represented therein. Casey and Crawford [20] defines MIR as strategies for enabling access to music collections, both new and historical, need to be developed to keep up with expectations of search and browse functionality.

These strategies are collectively called MIR and have been the subject of intensive research by an ever-increasing community of academic and industrial research laboratories [2]. For the purpose of this paper, we define MIR as the field or system concerned with the consistent developments of theories and techniques aiming at the smooth retrieval of music and musical information for the real-world applications. To further understand the definition, we explain the functionality of MIR in the subsequent section.

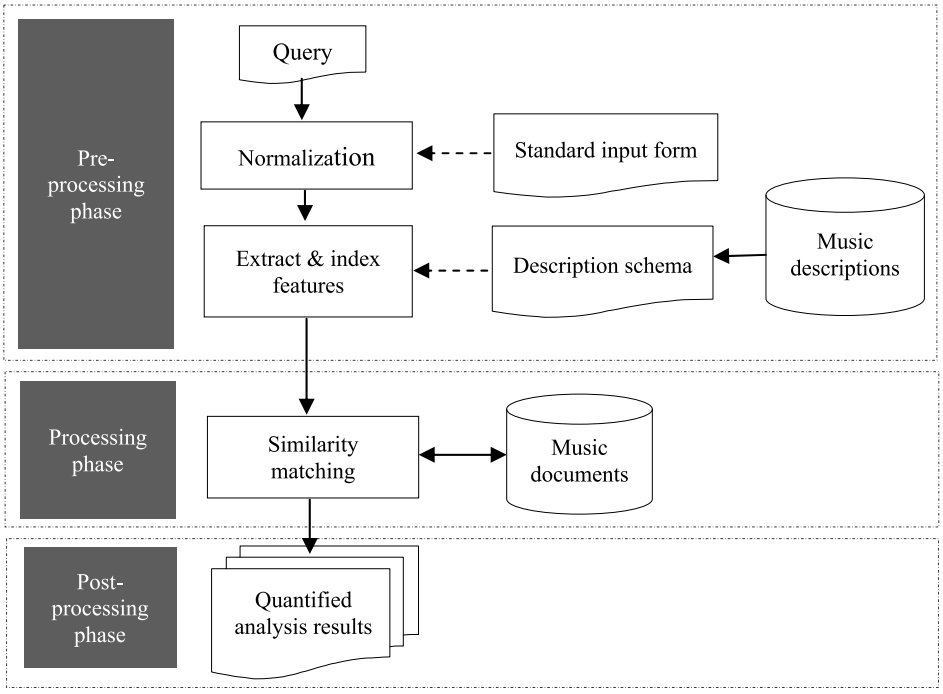

Fig. 1. Functionality of an MIR system.

In general, MIR systems have been built around three main phases: pre-processing phase, processing phase, and post-processing phase. Fig. 1 shows these three phases with their functions in an MIR

system. In the pre-processing phase, we acquire the required data and organize it for further uses. The sub-phases here are: feature extraction, data normalization, and data storage. Although feature extraction itself is a huge area of research in MIR, however, it falls under the pre-processing phase, where features such as pitch, timbre, loudness are extracted from audio files. Later in the data normalization and data storage phases, these extracted data are normalized according to the standard input form and then stored and indexed in the database. The processing phase performs similarity comparison of the features, which is the main focus of this paper. Here a combination of similarity comparison algorithms is used to compare the stored features. The post-processing phase which culminates in the parameter specific quantified search results as computed in the processing phase. Although each phase has its significance, two of the most challenging aspects of MIR systems are feature extraction and similarity comparison. Most of the research work has been done with a focus on these two aspects separately, and only less work has been done focusing on both aspects together. We thus discuss this in the following sub-sections in detail. In particular, since we focus on similarity comparison methods much more than feature extraction, we limit the feature extraction part to bellow mentioned brief introduction only.

3.1 Feature Extraction

As stated earlier we paper mainly focus on surveying the various similarity analysis methods of MIR data, but extraction of audio features is also an essential pre-requisite. So, here we provide a summary of extracting features from audio data. Music features such as pitch, loudness, duration, and timbre are critical for MIR because these features provide retrieval capabilities as per context or user needs and are also used to index large databases. All these features except timbre are objective by nature and can be modeled accurately. Timbre, however, is subjective and contains distinctive qualities which include amplitude envelope, harmonicity, and spectral envelope.

A typical retrieval application uses a combination of various methods to extract audio features along with the traditional text queries. Lambrou et al. [21], Zhang and Kuo [22], and Tzanetakis and Cook [23] are some of the primary researchers on feature extraction in MIR. Later Typke et al. [24] described commonly extracted features and how they are computed. Recently, Alien et al. [25] present taxonomy of feature extraction techniques into two main categories: physical techniques and perceptual techniques. They also present several subcategories under these two main categories. In Fig. 2, we briefly explain the general feature extraction process. We describe a step-by-step mechanism, based on a combination of approaches presented by Alias et al. [25] and Mitrovic et al. [26]. As the figure demonstrates, feature extraction operates in four phases: conceptual phase, pre-extraction phase, extraction phase, and post-extraction phase. The first step in the conceptual phase is to analyze the application-specific requirements. In the second step, feature designs are established, which refers to determining what aspects of an audio signal the features should capture [25]. This is performed in the context of the application domain and the specific retrieval task. The next phase is the pre-extraction phase, where the first step is to implement the designs and develop a technical solution to fulfill the specified requirements.

In the second step, we develop the segmentation of audio signals captured from the input device. This segmentation is performed with the help of window function, which first transforms these signals into shorter units and later converts them into a continuous signal of finite blocks. In the extraction phase,

first the audio features are extracted from the audio signals with the specific retrieval technical solution. Then these features are transformed into the required set up. The aim is to retrieve the features which provide a compact yet descriptive vision and carry the salient acoustic characteristics of a signal. The methods used for extracting these features depend on the physical or perceptual basis of the signal contents. Moreover, based on the application-specific requirements, time or frequency based information of these features is captured. To keep this information, the features extracted from subsequent blocks are merged into a single feature vector [25]. Lastly, in the post-extraction phase, we conduct an audio analysis task upon the feature vectors obtained in the previous step. The audio analysis works as a generic label to encompass any application-specific audio processing requirements. Alias et al. [25] and Mitrovic et al. [26] present a detailed taxonomy of features and techniques used in feature extraction. In Section 4.3, we provide an overview and classification of the features extracted for similarity measures.

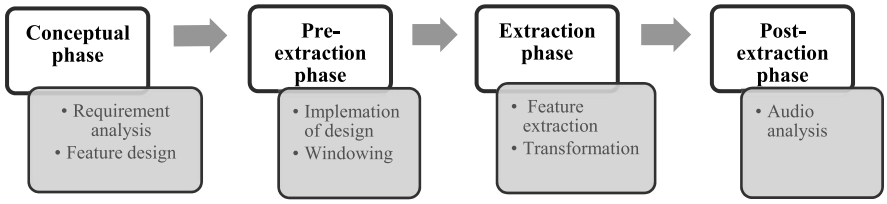

Fig. 2. Mechanism of feature extraction process.

3.2 Similarity Comparison Methods

The most fundamental goal of MIR systems is to retrieve music or musical information with the utmost accuracy, efficiency, and convenience for the end user. To attain this aim, one music piece is compared with another and based on the comparison a similarity result is produced. The similarity is quantified by assuming a similarity score, the more the similarity, the higher the score. A large number of efficient methods have been developed to compute the similarity of given musical data and assign a similarity score. In general, the similarity comparison methods are categorized into two broad categories: (1) note-based methods, where a set of strings or notes is compared for similarity scores and (2) frame-based methods, where a frame of musical data is compared using image-comparison methods. In addition, there are application specific approaches such as ground truth based approach, probability matching approach, and N-gram approach. These approaches are more concerned with the overall process of getting results rather than just comparing two music files. In this section we discuss and compare some of the major methods including string matching, geographic distance measurement, N-grams, ground truth, probability, and hybrid approaches.

String matching approach: String matching, in general, refers to comparing two given strings and finding the similarity between them. Before being introduced to music retrievals by Mongeau and Sankoff [8], string matching was successfully used in the fields like text processing, data compression and bioinformatics Measuring [27-29]. To apply string matching in music retrievals, a monophonic musical piece can be represented by one-dimensional strings of characters, where strings can represent pitch of the notes, interval sequences, gross counter, and etc. A musical piece can be described as sequences of integers, representing pitches of notes. The methods introduced for content-based music

retrieval on string matching have largely been applications of conventional approximate string matching approach (note-based methods). Here the sequences of notes containing information about pitches and respective durations of the music data are represented as a string of integers. This approach has been further refined for better efficiency by introducing an equivalent interval between two successive pitches instead of absolute pitch levels. On similar lines, matching becomes tempo invariant by quantifying the duration ratios between two consecutive notes. Moreover, by using the intervals between consecutive pitches, matching is made transposition invariant. The edit-distance approach introduced by Mongeau and Sankoff [8] presented the computation of sequences which are not same in the length. Later authors such as Gusfield [27] and Robine et al. [30] worked on the improvement of edit-distance methods. Some of the commonly used algorithms for string matching are edit distance [8,12,31,32], Levenshtein edit distance [33], dynamic programming algorithm [34], local similarity problem [35], and discrete time warping [36]. Simply calculating similarity does not work for strings with different lengths. So, to deal with the different lengths of strings suitable sub-strings is chosen. Mongeau and Sankoff [8] introduced the comparison of different length strings.

Geometric representation: Although the string matching approach works satisfactorily for monophonic music by representing by a linear sequence of pitches, it is not enough for polyphonic music, which has several transposed notes and themes. For similarity comparison and comparative analysis (viz. styles of different music composers) of polyphonic music, there was a need for a comprehensive model of representation and subsequent music retrieval algorithms. Geometric representation is the frame-based comparison where two frames are compared with each other similar to two images being compared, which works for both monophonic and polyphonic music. In the pre-processing phase, music data is represented as a frame containing strings in geometrical form. Unlike string matching methods geometric methods do not assume that notes are ordered. Dovey [37] and Ukkonen et al. [38] introduced the piano-roll representation of music data, which can also be categorized as geometric representation or frame-based methods. This concept of geometric representation was further refined in the form of a piano roll [38,39] and earth mover's distance (EMD), which is an efficient algorithm originally used for image retrieval systems [38,40]. Geometric representation proposes an extended representation of musical data, including the duration of notes as an additional feature. Recently, Bryan et al. [41] also proposed a comprehensive approach of source separation from polyphonic music representation (piano roll).

N-grams approach: N-grams approach introduced by Downie [11] uses the typical database query and retrieval mechanism for similarity comparisons. Musical n-grams are developed from polyphonic music using its pitch and rhythm dimensions. The method converts music data into discrete units of music information, as the words are discrete units of language. These discrete units of music information are called as "musical words". These words can be used directly in the traditional text based information retrieval methods as the "real words". Moreover, to evaluate the validity of this approach, several analyses were conducted to assess the matching of informatics properties between "musical words" and "real words". After the informatics analyses, a database of "musical words" was developed in accordance with the information retrieval system Further, this approach is also applied in text retrieval where the sequence of words remains as overlapped with constant-length subsequences [42]. The approach works successfully with both monophonic and polyphonic musical data. Further research in the field is taken forward by [42,43].

Ground truth-based approach: This approach focuses on the end results of similarity comparison. It compares a given musical set of data and forms an ordered list based on the rankings of the similarities to the established ground truth. A ground truth is established with the help of human experts and that ground truth is the basis for further comparisons and similarity rankings of the songs. The similarity between two music files is compared by finding out how close they are to the ground truth. The goal here is to not only find the matches to the query but also retrieve them in the right order. Considering this approach focuses on the results rather than the process, different similarity comparison methods are used in various researches. However, it would not be practically possible to establish a ground truth for millions of music files. So, the filtering mechanism is used to exclude music files which are very different from the query. The filtering is done by using SQL queries based on the features such as Pitch range, Duration ratio, etc. Subsequently, the groups of music files are created based on the similarity to the query. Applied to the database, the rankings of music files can be achieved in accordance with the ground truth. In addition, this ground truth serves as a basis for comparing the accuracy of MIR systems [24,39,44,45].

Probabilistic matching: The probability matching approach can also serve as an improvement over other similarity methods. The Markov model is very well known as the probability theory. Considering the large size of music databases, it suits the music retrieval process as well. Probability matching methods aim at the probabilistic aspects of candidate music pieces and then compare them with corresponding probabilities of queries. Basic music features such as pitch, interval, or rhythm are used for calculating Markov chains. Features like a certain pitch, note duration, or interval can correspond to states in the Markov chains. Representative in [46,47] are simple and implements first-order Markov chains for modeling the rhythmic and melodic contours of a music piece. Later Hoos et al. [48] proposed the GUIDO system, which uses Markov model, describing probabilities of state transition and compares the metrics of transition probabilities. The transition probabilities show the numbers of occurrence for the various subsequent states. To compute the similarity of a query music file in the database, we calculate the product of the transition [49].

Hybrid approach: The method proposed by Mullensiefen and Frieler [50] works as a combination of several methods, therefore it is called a hybrid approach. This approach uses the measures of three categories: vector measures, symbolic measures, and musical (mixed) measures to compute the similarity between music pieces. The vector approach is the one in which the music data is represented as vectors in a suitable real vector space, and other methods such as scalar products and means of correlation can also be applied here. The symbolic approach represents music data as strings. Lastly, the musical approach uses the musical knowledge as the key music data, and it can be represented as either vector, or symbolic, or as scoring models. The approach uses state-of-the-art techniques introduced in the field of MIRs for musical data transformation and similarity matching.

Table 1 presents a brief comparative description of the similarity comparison methods explained so far. In addition, there are some new methods introduced by Foster et al. [51] and Park and Lee [52]. However, the methods of melodic data representation and similarity measurement employed in this study continues to work as foundation of the further researches. Moreover, systemizing and analyzing these approaches led to the idea of constructing several new similarity measures such as Mullensiefen and Frieler [53]. To further elaborate MIR systems and similarity measurements, in particular, we elaborate the components of MIR systems in the next section.

Table 1. Comparative description of similarity measures

Similarity measures	Advantages	Limitations
String matching	Easy to implement and develop further. Improved accuracy with newly developed researches.	Time consuming Work only with Monophonic music. No consideration of duration of notes.
Ground truth	Provide a common ground for the comparisons and rankings of various MIR systems	No qualification as a similarity measurement method. Increased complexity with establishing ground truth.
N-grams	Easy to implement and time efficient Works for both monophonic and polyphonic music.	Limited to database mechanism. No inclusion of musical technicalities.
Geometric	Time efficient. Work with monophonic and polyphonic music. No consideration of notes duration.	Accuracy is limited. Complex implementation.
Probabilistic approach	Time efficient. Suitable for large data bases. Can also be applied to other MIR systems for faster computations.	Limited accuracy. Time consuming implementation.
Hybrid approach	Application specific. Can use variety of approaches as one, improvising on speed and efficiency.	Complex implementation Need to be developed on case by case basis.

4. Components of MIR Systems

We covered the functionality of MIR systems in the previous section. In furtherance of the understanding of MIR systems in detail, we now describe their main components this section. As mentioned before, this paper focuses on similarity comparison methods, and thus, the components described in the section including algorithms, features used for comparison, and type of music data, are mainly related to similarity comparisons.

4.1 Algorithms for Distance Measures

In the pre-processing phase, specific features are extracted from the given audio files. After then, these features are transformed into standardized templates suitable for comparisons. In the processing phase, one or more combinations of several similarity comparisons algorithms are applied to these templates to evaluate the similarity of two given music pieces. The goal here is to introduce the commonly used algorithms. The traditional way of sequence comparison involves comparing strings of equal lengths by comparing the positional values of the both strings. This traditional approach got further refined into a modern theory of finite sequence comparison which started the development of distance metrics into the music comparison [29]. The geometric representation, such as piano roll represents a piece of music as a collection of horizontal line segments or points in the two-dimensional Euclidian space. The horizontal axis represents the time, and the vertical represents the pitch values. Moreover, the occurrence of a point set is searched within another point configured to find the match [38]. Several distance metrics have been implemented ever since some of them are discussed as below.

Edit distance: Edit distance or Levenshtein distance [33] is the most commonly used distance algorithm for music sequences. For quantifying the difference between two strings, it is to count the minimum required transformations to transfer one string into another. Later Wagner and Fischer [34] proposed advancements with dynamic programming, which solves the problem of two strings with different lengths.

Smith-Waterman: Smith-Waterman algorithm [35] also known as the local alignment [27] is used for measuring local similarity of strings. The algorithm performs local sequence alignment for determining similar regions between two strings. So, instead of looking at the total sequence, the algorithm compares all the segments of possible lengths to optimize the similarity measure. The algorithm is more suitable for finding a substring with the highest similarity by just focusing on local alignment rather than global one [54].

Earth mover's distance: This distance is basically used for image retrieval [40], but Typke et al. [24] employed it efficiently for similarity measurement of musical data. The algorithm is used in both string matching and geographic representation approaches. A note represents a point between two-dimensional space of time and pitch in MIRs. Here the weight of that note is represented by its duration and its position. However, EMD is not preferred because of its high computing cost in comparison to traditional methods like Euclidean distance [55].

Dynamic Time Warping (DTW): DTW is a robust distance algorithm mainly used for time-series or image retrieval. However, music is represented as a 2-dimental time-series of pitch and note duration. DTW discovers an optimal alignment between two given time series, and computes the matching cost corresponding to that alignment [56].

Hybrid or adapted algorithms: There are several algorithms which were adopted as original but slightly changed as per requirements of the field. Here are some examples. Robine et al. [30] presented improvements in Edit-based algorithms as per music theory. Ferraro and Hanna [54] presented improvements in local alignment algorithms introduced by Smith and Waterman [35]. Foster et al. [51] use pair-wise Euclidean distance. Park and Lee [52] used combinational Euclidean distance.

Some of MIR systems prefer only exact matches or cases where the search string is a substring of the database entries. For such uses, standard string searching algorithms such as Knuth-Morris-Pratt and Boyer-Moore are used. Also, standard text indexing methods such as B-trees and inverted files are used for finding substrings that match exactly [49]. Some of the other commonly used algorithms in the field of MIRs are Euclidean distance, Hamming distance, cosine similarity, etc.

4.2 Music Types used in MIR

An important challenge of the MIR system is to deal with the different music types. The MIR system needs to operate with monophonic, polyphonic, and sometimes both type of musical data. The sort of music data used depends on the specific application area of the MIR. In this section, we present a brief idea about monophonic and polyphonic music representations. Most MIR systems work with either monophonic or polyphonic music data. In simple language, monophonic refers to one note at a time while polyphonic refers to two or more notes at a time. For example, a piano playing a melody would be monophonic music while two or more instruments playing a melody at the same time would be polyphonic.

Monophonic music is represented by a one-dimensional string of characters where every character represents one note of consecutive notes. Traditional string matching algorithms such as editing distances, finding occurrences of one string in another has been applied for monophonic music data. Some of the examples of monophonic music data are in [8,30,52].

Polyphonic music, on the other hand, is represented with two or more strings being played simultaneously. However, traditional string matching methods are not suitable for the polyphonic music data. Therefore, frame based methods (geographical methods), N-gram approaches, and other hybrid methods are used for polyphonic music. Refer to [39,57,58] as examples of polyphonic music data. Although the polyphonic methods can be used for both monophonic and polyphonic music types, they are dealt separately with application specific goals.

4.3 Features Considered for Comparison

Before comparing two musical pieces, an MIR system must extract quantitative features from audio files. In this section, we briefly present some of the features used for defining similarity measures. Categorizing audio features is challenging due to their manifold nature. Hence, they are classified based on their applicability. Tzanetakis [59] proposes two principles for the taxonomy of the features. The first principle is related to the computational issues of features, and the second principle corresponds to the qualities like texture, timbre, and pitch. Peeters [60] presents four principles for the feature taxonomy steadiness, time-extent, abstractness, and extraction process of features. Mitrovic et al. [26] proposes a method-oriented approach which is based on the internal structure and similarities of the features. Moreover, based on this approach mentioned above, Fig. 3 shows the categorization of features presented by Mitrovic et al. [26].

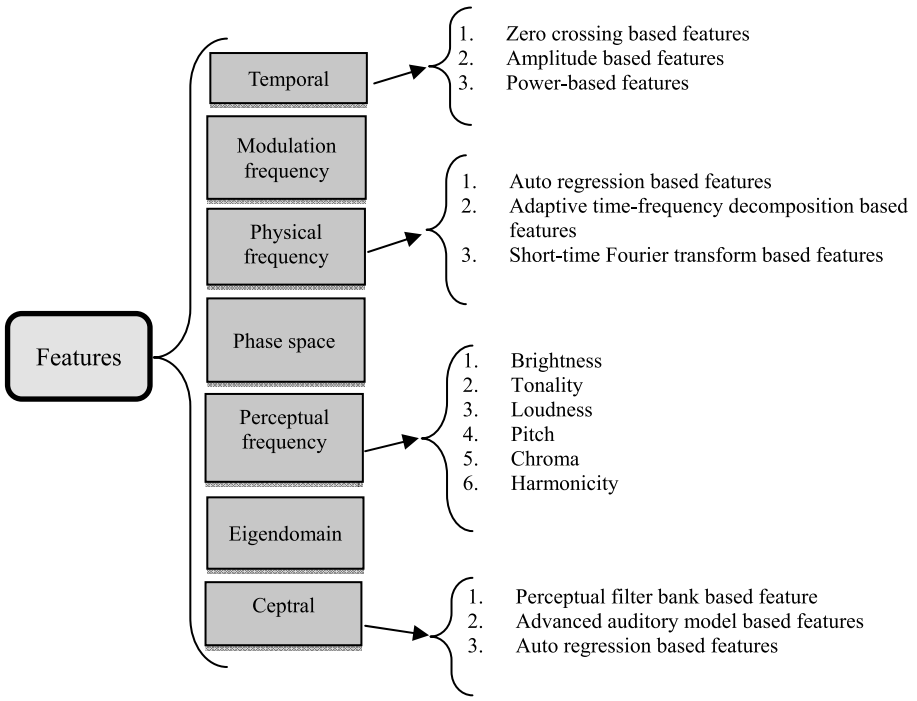

Fig. 3. Tree based representation of musical features.

Audio feature extraction is a huge area of MIR systems, for further study of the audio features, we recommend [25] and [26]. We have explained components of MIR systems in general in this section. However, the components can differ based on the application area of a particular MIR system. In the next section, we present the major application areas of the field.

5. Application of MIR Systems

With music becoming digitally available, the application areas of MIR have also grown in range. In this section, we present some of the major application areas of MIR systems. We classify the application areas into the following eight categories: music retrieval, copying detection, music recommendation systems, automatic playlist generation, music genre classification, music education and training, popularity estimation and, others.

Music retrieval: Music retrieval aims at helping end-users search and finds a desired piece of music from an extensive database. Music retrieval itself is a huge application area, and it is further categorized into two basic categories. The first is the semantic-based or category-based retrieval, and the second is the query by example retrieval. In semantic retrieval, the user searches for a semantic or a tag such as happy, sad songs, rock, and metal music [61,62] or find me all songs with C major [63]. Laurier et al. [64] presented a method for automatic music mood recognition. Kim et al. [65] analyzed several "music mood" recognition methods in his review article. Most of the commercial music applications such as SearchSounds1 and Gedoodle2 use semantic or category-based retrievals. Semantic retrievals such as music genre classifications are explained in the subsections. Query by example is based on comparing a query music piece against a database. We elaborate a query through example applications by further subcategorizing them into two categories: audio recognition and audio alignment.

- Audio recognition's goal is to identify or retrieve a queried piece of music against a database. There are some commercial applications such as Shazam and Gracenote which provide queried musical pieces as results for the end user.

  These applications are based on the approach proposed by Wang [66]. Audio recognition also includes techniques such as audio fingerprinting, query by humming, and cover song identification. Audio fingerprinting is used in different commercial applications for identifying music and different sounds [67]. Query by humming is the retrieval of music based on melodic inputs hummed by a user as a query to the database. Dannenberg et al. [68] proposed an approach called MUSART based on query by humming retrieval. Recently, Salamon et al. [69] proposed a version identification method adopted by the query by humming approach. Cover song identification is the retrieval of all versions of the same song. These versions may differ from the original in instrumentation, key, tempo, and the language of the vocals or structure [70]. There are several commercial applications such the musical covers and cover project available for the study of musical influences.
- Audio alignment refers to the task of synchronizing two audio sequences with similar musical content in time. In addition to identifying the given music piece, the goal is to link time positions. In process of being mapped with the time points, a music files must be analyzed. In musicology,

the information is used to analyze several recordings of the same piece of music. In film productions, auto synchronization helps match the re-recorded track to the original studio recording [71]. Audio alignment enables quick access to a particular part of the recording which makes it easier to compare with other parts. Research works such as MATCH by Dixon and Widmer [72], and system by Muller et al. [73] are applying variants of the dynamic time warping algorithm.

Copying detection: The revenue loss due to music plagiarism has grown drastically over the years. In South Korea alone, it was more than 2 billion in the year 2009 [74]. There are several organizations such as ASCAP (American Society of Composers, Authors, and Publishers) or SACEM (Society of Authors, Composers, and Publishers) which work for the copyrights and licensing of the music industry. The goal here is to identify whether a song or part of it has been copied from another song or not. For the same purpose, similarity between music pieces are computed, and a score is generated to measure it. There are many applications (famously called audio forensics toolbox) commercially available which work as plagiarism detectors, for example, music plagiarism analyzer [75] and Fraunhofer. Some of the approaches are Park et al. [76], Lee et al. [74], and Saracoglu et al. [77].

Music recommendation system: This system proposes a list of music pieces based on the listening behavior and rating history of a user. A good recommendation system requires these core qualities, accuracy, diversity, transparency, and serendipity [78]. The two most commonly used methods for recommendation systems are collaborative filtering algorithm [79] and contend-based model. If we combine both the methods, a list of features such as pitch, rhythm or genre can be generated [80]. In general, music recommendation systems consist of three components. First, user modeling includes information about the user such as age, gender, region, and mood. Second, music data profiling, which refers to meta data such as editorial, cultural and acoustic. Third, query type refers to the type of queries [80]. There are some well-known commercial systems in existence such as Last.fm and Pandora. Several other methods have been proposed in recent years with the improved focus on user-aware, personalized, and multi-model recommendations. For instance, Baltrunas et al. [81] proposed a system called In-Car Music, Zang et al. [82] presented the Auralist music recommendation system, and Schedl and his colleagues [83,84] presented location-aware music recommendation systems [77].

Music genre classification: The aim of music genre classification is categorizing and labeling the music into genres for improved management of music. This classification helps in arranging the music at stores and on the web. The classification also helps users to form likability and dis-likability towards a particular type of music. The musical genre is characterized by the properties such as rhythmic structure, instrumentation, and harmonic context of music. Before the proliferation of digital music, genre classification had been performed manually, as humans are remarkably good at this feature [85]. As shown in Fig. 4, music can be categorized into several categories and subcategories. Researchers such as [23], [86], [87], and [88] suggested improvements with the help of research work in this field.

Automatic music playlist generation: This application refers to generating a list of ordered songs based on previous choices and behaviors of the end user. Although playlist generators are very similar to music recommendation systems, there are two basic differences between the systems. First, a recommendation system proposes new tracks based on the history of the user while the playlist generator recognizes the known tracks. Second, a playlist generator does not provide an ordered list,

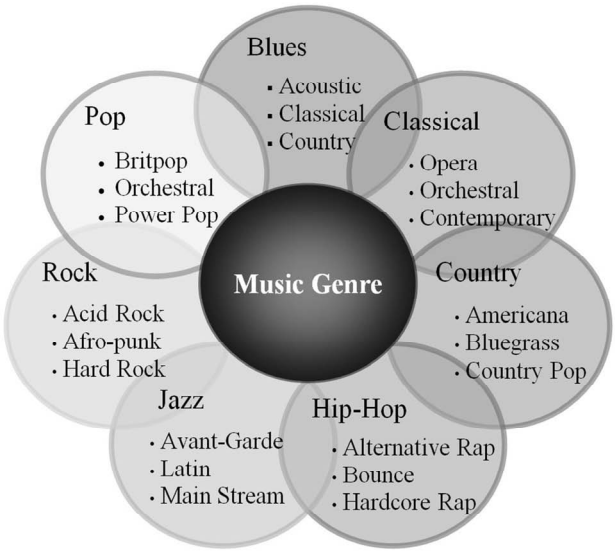

Fig. 4. Music genre classification.

whereas a music recommendation system does. According to a study conducted by Pohleet et al. [89], the following tracks should be similar to form a good playlist. But continued similarity in music can feel monotonous to the user [78]. Later Schedl et al. [83] defined several requirements of a playlist generation system in addition to the similarity, which is the popularity of the artist, the trendiness of the artist, recentness of the track, and the novelty of the artist and track. The existing commercial examples of these systems can be Intelligent iPod7 [90] and YAMAHA BODiBEAT [78].

Popularity estimation: Popularity estimation aims at predicting whether a given music piece is likely to be popular before its public release. All parties involved in the music industry get greater help by being able to predict the hits of the future. There are similar approaches focused on the rankings of songs on popularity basis by Koenigston and Shavitt [91] or predicting artists that are popular in a particular area by Schedl et al. [92]. Dhanraj and Logan [93] propose an interesting approach, which uses both acoustic and lyrical information to analyze the pattern. Karydis et al. [94] introduce the Track Popularity Dataset (TPD), which is a collection of track popularity data.

Music education and training: Music retrieval techniques can be used in understanding the music better. A possible example here is the field of computational music theory. It describes an area where music content description techniques offer study related to music such as music pieces, analysis of previously composed music and comparative studies using large datasets. It also formalizes expert knowledge as well as building and maintaining digital music libraries for research issues involving music retrieval and training. For example, Serra et al. [95] presented a detailed study of MIR areas to extend the context of music information research area. Also, there are some commercial applications such as HumOn, which records the hummed voice and changes it into various composed musical forms. The user can analyze this sound, classify the notes, and play it with different music genres.

Other application: There are MIR application areas which are not limited to standard retrieval scenarios. The areas mentioned below are not described in detail, however they have considerable importance in the musical field. Following areas that fall in above mentioned category.

- Solving crimes or helping entertainment industry with different sounds.
- Automatic music composition, for example, IBM's Watson.
- Track separation and instrument recognition.
- Automatic music transcription (audio recording into symbolic notations).
- Music listening experiences, to improve mobile music recommendation.
- Pattern spotting.
- Audio mosaicking, where a target music track is analyzed and its audio descriptors are extracted for small fragments.

As MIR is a growing research field, the application areas are also growing. An article by Serra et al. [95] is recommended for the further study of MIR applications. Moreover, we summarize the reference details of the MIR application areas in Table 2.

Table 2. Application areas of MIR

Applications	Ref.
Semantic or category-based retrieval, find music with tags, mood recognition	[64], [66], [68], [69], [70], [71], [72], [73], [96]
Artist or instrument recognition	[65]
Audio fingerprinting	[67], [97]
Query by humming	[68], [69], [98], [99]
Cover song identification	[70], [100]
Copying detection	[74], [75], [76], [77]
Music recommendations	[78], [79], [80], [81], [82], [83]
Playlist generations systems	[78], [84], [86], [90]
Music genre classification	[23], [61]
Music education and training	[101], [102]
Popularity estimation	[91], [92], [103]
Others	[101], [102]

6. Comparative Analysis of MIR Systems

We presented a comparison of MIR systems based on the algorithms used in Section 4 and in this section we compare these systems based on their accuracy and efficiency. However, the comparison of MIR systems is challenging for the following reasons. (1) MIR systems are application specific, and they are developed keeping a particular task in mind. Some applications are focusing on the accuracy while others are focusing on speed of retrieval. (2) Features considered for the similarity comparison can be different for two different MIR systems. (3) The overall architecture or the some of the components of different systems can be different, so comparing these systems as a whole is not appropriate. However, for the purpose of evaluation, we compare and analyze some of the MIR systems based on accuracy and efficiency in this section.

6.1 Accuracy Based Comparison

Measuring accuracy of an MIR system is a tricky task as most of the approaches are developed for a

specific application area. For example, an approach developed for forensics, which compares and detects the similarity of sounds to solve any particular crime, is required to focus on accuracy. But, a general music listening and retrieval application usually focuses on both speed and accuracy of retrieval. Also, the evaluation mechanisms used by some of the accuracy analyzers such MIREX [104] handle MIR systems as black boxes, which measure their accuracy as a whole. Therefore, it disregards the effectiveness of a particular component of the system. Moreover, the computation time is usually not taken into account in the evaluation process [105]. Scholz et al. [106] present a quantitative index to help understand bottlenecks of evolution or measurement issues of MIREX. However, to present a comparative analysis, we present a table proposed by Robine et al. [30]. Here, average dynamic recall (ADR) explained in [39] is the specific measure proposed for accuracy computation. In addition to that, a ground truth established with the help of music experts works as the basis for the comparison. Table 3, explains the rankings of MIR systems. The traditional edit distance [107] is at the bottom of the table whereas the improved edit-distance [30] leads the accuracy table. For current comparisons, we recommend MIREX, an association dedicated to, various analysis of MIR systems.

Table 3. Accuracy results from Robine et al. [30] during MIREX 2005

Algorithm	Author	Average ADR
Improved edit-distance	Robine et al. [30]	77
Edit distance I/R	Grachten et al. [108]	66
N-grams	Orio [109]	65
Simple N-grams	Uitdenbogerd and Zobel [58]	64
Geometric	Typke et al. [39]	57
Geometric	Ukkonen et al. [38]	56
Edit distance	Lemstrom and Tarhio [107]	54

6.2 Efficiency Based Comparison

The methods developed for one type of application do not work efficiently with the other. Therefore, the efficiency of the methods can be compared if we do not treat the method as a whole and set certain criteria for it. We follow these criteria to proceed further in efficiency comparison: computation cost, music type, simplicity and accuracy.

For computing cost, there are applications such as music retrieval where a piece of music is compared with a large database having millions of songs. Therefore, the speed of the comparison becomes critical criteria. In general, the frame-based (image processing) approaches should be faster than the string-based approaches. However, the adjustments frame-based approach made to deal with polyphonic music makes the comparison almost negligible. In experiments conducted with simple edit-distance and local-alignment edit distance, the simpler one works slightly quicker. Moreover, the approaches with advanced algorithms such DTW are slightly faster than them too. For music type, as explained in Section 5, some of the MIR systems such as the works [8,30] with only monophonic music data whereas other methods such as the work [107] for both mono and polyphonic music data. Simplicity is an important part of an MIR system as it affects the understanding of a system, which affects further developments of the system. In general, frame-based approaches comparing two windows of polyphonic musical data are more complex than a simple edit-based approach. Regarding accuracy,

Table 3 explained in a previous section describes a comparative analysis of the accuracy of MIR systems. For further details, we recommend Typke et al. [24] as a representative survey which provides an analysis of the efficiency of commercially available applications.

7. Future Challenges and Summary

In this section, we present the current and future challenges of MIR and summarize the paper. MIR is a newly established multidisciplinary field; in fact, the first conference of the field was conducted in 2000 [110]. Despite the field being established, there are still many challenges, and we discuss some of them in this section.

Data availability, the availability of data, particularly for the less famous music is one of the challenges MIR research area is facing. Lamere [111] mentioned the lack of assigned tags explaining the difficulty of finding detailed data. Besides, the websites and blogs can lead to false and noisy data. As stated by Jha [112], it takes 20–30 minutes per track of one expert's time to enter the metadata, so the cost of MIR development is also a major challenge. Practical barriers, another challenge MIR research field faces is the practicality of the developed techniques. Due to the increased size of resources and computational power, these techniques should be able to adapt and scale to the commercial sizes. Evaluation of the MIR area, the field should be able to take into account the validity of algorithms regarding the development of real-world applications. Moreover, the research works should pay attention to the end user evaluation. Integrating the existing resources, the MIR research field should be able to the existing diverse multimodal Features and integrate them to build an improved MIR system. Moreover, this improved version should be low in cost and benefit the end user. User-centered development, the MIR research area should be more focused on improving the overall user experience towards music. Which includes the listening of music, composing music, and studying music. Serra et al. [70] propose that the music research should capture social and cultural aspects of a given region.

Other research works have also presented the challenges of the MIR field. Downie [1] described some of the challenges as follows. (1) The research towards the further study to understand music should be encouraged. (2) The researcher should learn insights about music technicalities. (3) The research area should cover both musical ranges from modern to old classic. (4) We need to find the right balance between symbolic and metadata research and should develop a full-featured, robust, multifaceted system. Goto [113] mentioned challenges as follows: (1) delivering content-aware music for an individual, (2) predicting trends of music, (3) enriching human-music relationships by redefining authenticity of the music, (4) enhancing the listening experience of the end-user, and (5) contributing towards addressing the environmental issues. Serra et al. [70] proposed that research works should be focused on technological, user social-cultural, and exploitation perspective of music.

In summary, the MIR research area is newly established, and the opportunities of the field are vast. In this paper, we reviewed some the major MIR systems by introducing their history and basic functionality. We also presented a comparative analysis of some of the major music similarity comparison methods. In components of MIR, we explained and compared the algorithms, data type, music type used in MIR. Also, we presented the application areas of MIR systems. Lastly, we presented some of the challenges of the MIR research area. The paper helps in getting insights about the challenges and opportunities of the MIR research area.

Acknowledgement

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. R7117-17-0214, Development of an Intelligent Sampling and Filtering Techniques for Purifying Data Streams).

References

[1] J. S. Downie, "Music information retrieval," *Annual Review of Information Science and Technology*, vol. 37, no. 1, pp. 295-340, 2003.

[2] M. A. Casey, R. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney, "Content-based music information retrieval: current directions and future challenges," *Proceedings of the IEEE*, vol. 96, no. 4, pp. 668-696, 2008.

[3] M. Kassler, "Toward musical information retrieval," *Perspectives of New Music*, vol. 4, no. 2, pp. 59-67, 1966.

[4] M. Kassler, "MIR: a simple programming language for musical information retrieval," in *The Computer and Music*. Ithaca, NY: Cornell University Press, 1970, pp. 299-327.

[5] H. B. Lincoln, "Some criteria and techniques for developing computerized thematic indices," in *Elektronische Datenverarbeitung in der Musikwissenschaft*. Regensburg, Germany: Gustave Bosse Verlag, 1967, pp. 57-62.

[6] D. Deutsch, "Octave generalization and tune recognition," *Perception & Psychophysics*, vol. 11, no. 6, pp. 411-412, 1972.

[7] D. A. Byrd, "Music notation by computer," Ph.D. dissertation, Indiana University, Ann Arbor, MI, 1984.

[8] M. Mongeau and D. Sankoff, "Comparison of musical sequences," *Computers and the Humanities*, vol. 24, no. 3, pp. 161-175, 1990.

[9] J. Foote, "An overview of audio information retrieval," *Multimedia Systems*, vol. 7, no. 1, pp. 2-10, 1999.

[10] D. Byrd and T. Crowford, "Problems of music information retrieval in the real world," *Information Processing & Management*, vol. 38, no. 2, pp. 249-272, 2002.

[11] J. S. Downie, "Evaluating a simple approach to music information retrieval: conceiving melodic n-grams as text," Ph.D. dissertation, University of Western Ontario, London, Canada, 1999.

[12] A. Ghias, J. Logan, D. Chamberlin, and B. C. Smith, "Query by humming: musical information retrieval in an audio database," in *Proceedings of the 3rd ACM International Conference on Multimedia*, San Francisco, CA, 1995, pp. 231-236.

[13] A. Freed, "Music metadata quality: a multiyear case study using the music of Skip James," in *Audio Engineering Society Convention 121*. New York, NY: Audio Engineering Society, 2006, pp. 1314-1325.

[14] M. C. Jones, J. S. Downie, and A. F. Ehmann, "Human similarity judgments: implications for the design of formal evaluations," in *Proceedings of 8th International Conference on Music Information Retrieval (ISMIR 2007)*, Vienna, Austria, 2007, pp. 539-542.

[15] M. Schedl, A. Flexer, and J. Urbano, "The neglected user in music information retrieval research," *Journal of Intelligent Information Systems*, vol. 41, no. 3, pp. 523-539, 2013.

[16] J. Urbano, J. Morato, M. Marrero, and D. Martin, "Crowdsourcing preference judgments for evaluation of music similarity tasks," in *Proceedings of the 1st ACM SIGIR Workshop on Crowdsourcing for Search Evaluation*, Geneva, Switzerland, 2010, pp. 9-16.

[17] U. M. Julian, "Evaluation in audio music similarity," Ph.D. dissertation, Charles III University of Madrid, Madrid, Spain, 2013.

[18] J. S. Downie, "The scientific evaluation of music information retrieval systems: foundations and future," *Computer Music Journal*, vol. 28, no. 2, pp. 12-23, 2004.

[19] R. J. Demopoulos and M. J. Katchabaw, "Music information retrieval: a survey of issues and approaches," Department of Computer Science, University of Western Ontario, London, Canada, *Technical Report #677*, 2007.

[20] M. Casey and T. Crawford, "Automatic location and measurement of ornaments in audio recordings," in *Proceedings of 5th International Conference on Music Information Retrieval (ISMIR 2004)*, Barcelona, Spain, 2004, pp. 311-317.

[21] T. Lambrou, P. Kudumakis, R. Speller, M. Sandler, and A. Linney, "Classification of audio signals using statistical features on time and wavelet transform domains," in *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, Seattle, WA, 1998, pp. 3621-3624.

[22] T. Zhang and C. C. J. Kuo, "Audio content analysis for online audiovisual data segmentation and classification," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, pp. 441-457, 2001.

[23] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 5, pp. 293-302, 2002.

[24] R. Typke, M. den Hoed, J. de Nooijer, F. Wiering, and R. C. Veltkamp, "A ground truth for half a million musical incipits," in *Proceedings of the 5th Dutch-Belgian Information Retrieval Workshop*, Utrecht, The Netherlands, 2005, pp. 63-70.

[25] F. Alias, J. C. Socoro, and X. Sevillano, "A review of physical and perceptual feature extraction techniques for speech, music and environmental sounds," *Applied Sciences*, vol. 6, no. 5, article no. 143, 2016.

[26] D. Mitrovic, M. Zeppezauer and C. Breiteneder, "Feature of content-based retrieval systems" in *Advances in Computers: Improving the Web*. London, UK: Academic, 2010, pp. 71-150.

[27] D. Gusfield, *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*. Cambridge, UK: Cambridge University Press, 1997, pp. 91-95.

[28] S. Needleman and C. D. Wunsch, "A general method applicable to the search for similarities in the amino acid sequence of two proteins," *Journal of Molecular Biology*, vol. 48, no. 3, pp. 443-453, 1970.

[29] D. Sankoff and J. B. Kruskal, *Time Warps, Strings Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Reading, MA: Addison-Wesley, 1983.

[30] M. Robine, P. Hanna, and P. Ferraro, "Music similarity: improvements of edit-based algorithms by considering music theory," in *Proceedings of the International Workshop on Multimedia Information Retrieval*, Bavaria, Germany, 2007, pp. 135-142.

[31] K. Lemstrom, P. Laine, and S. Perttu, "Using relative interval slope in music information retrieval," in *Proceedings of the International Conference on Computer Music*, Beijing, China, 1999, pp. 317-320.

[32] R. J. McNab, L. A. Smith, I. H. Witten, C. I. Henderson, and S. J. Cunningham, "Towards the digital music library: tune retrieval from acoustic input," in *Proceedings of the first ACM International Conference on Digital Libraries*, Bethesda, MD, 1996, pp. 11-18.

[33] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707-710, 1966.

[34] R. A. Wagner and M. J. Fischer, "The string-to-string correction problem," *Journal of the ACM*, vol. 21, no. 1, pp. 168-173, 1974.

[35] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," *Journal of Molecular Biology*, vol. 147, no.1, pp. 195-197, 1981.

[36] Y. Zhu and D. Shasha, "Warping indexes with envelope transforms for query by humming," in *Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data*, San Diego, CA, 2003, pp. 181-192.

[37] M. J. Dovey, "An algorithm for locating polyphonic phrases within a polyphonic musical piece," in *Proceedings of the Artificial Intelligence and Simulation of Behavior (AISB)'99 Symposium on Musical Creativity*, Edinburgh, UK, 1999, pp. 48-53.

[38] E. Ukkonen, K. Lemstrom, and V. Makinen, "Geometric algorithms for transposition invariant content-based music retrieval," in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, Baltimore, MD, 2003, pp. 193-199.

[39] R. Typke, R. C. Veltkamp, and F. Wiering, "Searching notated polyphonic music using transportation distances," in *Proceedings of the 12th Annual ACM International Conference on Multimedia*, New York, NY, 2004, pp. 128-135.

[40] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *International Journal of Computer Vision*, vol. 40, no. 2, pp. 99-121, 2001.

[41] N. J. Bryan, G. J. Mysore, and G. Wang, "Source separation of polyphonic music with interactive user-feedback on a piano roll display," in *Proceedings of the 14th International Conference on Music Information Retrieval (ISMIR 2013)*, Curitiba, Brazil, 2013, pp. 119-124.

[42] S. Doraisamy and S. Ruger, "Robust polyphonic music retrieval with n-grams," *Journal of Intelligent Information systems*, vol. 21, no. 1, pp. 53-70, 2003.

[43] S. Downie and M. Nelson "Evaluation of a simple and effective music information retrieval method," in *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, 2000, pp. 73-80.

[44] R. Typke, P. Giannopoulos, R. C. Veltkamp, F. Wiering, and R. van Oostrum, "Using transportation distances for measuring melodic similarity," in *Proceedings of the 4th International Conference on Music Information Retrieval (ISMIR 2003)*, Baltimore, MD, 2003, pp. 193-199.

[45] R. Typke, R. C. Veltkamp, and F. Wiering, "A measure for evaluating retrieval techniques based on partially ordered ground truth lists," in *Proceedings of 2006 IEEE International Conference on Multimedia and Expo*, Toronto, Canada, 2006, pp. 1793-1796.

[46] C. Dodge and T. Jerse, *Computer Music: Synthesis, Composition, and Performance*. New York, NY: Schirmer Books, 1997, pp. 361-368.

[47] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*. Oxford, UK: Oxford University Press, 2001, pp. 245-288.

[48] H. H. Hoos, K. Renz, and M. Gorg, "GUIDO/MIR–an experimental musical information retrieval system based on GUIDO music notation," in *Proceedings of the 2nd International Conference on Music Information Retrieval (ISMIR 2001)*, Bloomington, IN, 2001, pp. 41-50.

[49] R. C. Veltkamp, F. Wiering, and R. Typke, "Content based music retrieval," in *Encyclopedia of Multimedia*, Boston, MA: Springer, 2008, pp. 97-98.

[50] D. Mullensiefen and K. Frieler, "Melodic similarity: approaches and applications," in *Proceedings of the 8th International Conference on Music Perception and Cognition*, Evanston, IL, 2004.

[51] P. Foster, M. Mauch, and S. Dixon, "Sequential complexity as a descriptor for musical similarity," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 1965-1977, 2014.

[52] M. W. Park and E. C. Lee, "Similarity measurement method between two songs by using the conditional euclidean distance," *WSEAS Transactions on Information Science and Applications*, vol. 10, no. 12, pp. 381-388, 2013.

[53] D. Mullensiefen and K. Frieler, "Cognitive adequacy in the measurement of melodic similarity: algorithmic vs. human judgments," *Computing in Musicology*, vol. 13, no. 2003, pp. 147-176, 2004.

[54] P. Ferraro and P. Hanna, "Optimizations of local edition for evaluating similarity between monophonic musical sequences," in *Proceeding RIAO 2007 Large Scale Semantic Access to Content (Text, Image, Video, and Sound)*, Pittsburgh, PA, 2007, pp. 64-69.

[55] Y. Tang, U. Leong Hou, Y. Cai, N. Mamoulis, and R. Cheng, "Earth mover's distance based similarity search at scale," *Proceedings of the VLDB Endowment*, vol. 7, no. 4, pp. 313-324, 2013.

[56] J. Lijffijt, P. Papapetrou, J. Hollmen, and V. Athitsos, "Benchmarking dynamic time warping for music retrieval," in *Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments*, Samos, Greece, 2010, pp. 1-7.

[57] S. Doraisamy, "Polyphonic music retrieval: the N-gram approach," Ph.D. dissertation, University of London, London, UK, 2004.

[58] A. Uitdenbogered and J. Zobel, "Melodic matching techniques for large music databases," in *Proceedings of the 7th ACM international conference on Multimedia*, Orlando, FL, 1999, pp. 56-66.

[59] G. Tzanetakis, "Manipulation, analysis and retrieval systems for audio signals," Ph.D. dissertation, Princeton University, Princeton, NJ, 2002.

[60] G. Peeters, "A large set of audio set for sound description (similarity and classification) in the CUIDADO project," 2004 [Online]. Available: http://www.citeulike.org/group/1854/article/1562527.

[61] P. Knees, T. Pohle, M. Schedl, and G. Widmer, "A music search engine built upon audio-based and web-based similarity measures," in *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Amsterdam, The Netherlands, 2007, pp. 447-554.

[62] D. Turnbull, L. Barrington, M. Yazdani, and G. Lanckriet, "Combining audio content and social context for semantic music discovery," in *Proceedings of the 32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Boston, MA, 2009, pp. 387-394.

[63] E. J. Isaacson, "Music IR for music theory," in *Proceedings of the Workshop on the Creation of Standardized Test Collections, Tasks, and Metrics for MIR and MDL Evaluation*, Portland, OR, 2002, pp. 23-26.

[64] C. Laurier, O. Meyers, J. Serra, M. Blech, P. Herrera, and X. Serra, "Indexing music by mood: design and integration of an automatic content-based annotator," *Multimedia Tools and Applications*, vol. 48, no. 1, pp. 161-184, 2010.

[65] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. J. Scott, J. A. Speck, and D. Turnbull, "Music emotion recognition: a state of the art review," in *Proceedings of the 11th International Conference on Music Information Retrieval (ISMIR 2010)*, Utrecht, The Netherlands, 2010, pp. 255-266.

[66] A. Wang, "An industrial strength audio search algorithm," in *Proceedings of 4th International Conference on Music Information Retrieval (ISMIR 2013)*, Baltimore, MD, 2003, pp. 7-13.

[67] C. S. Dhir, J. Lee, and S. Y. Lee, "Extraction of independent discriminant features for data with asymmetric distribution," *Knowledge and Information Systems*, vol. 30, no. 2, pp. 359-375, 2012.

[68] R. B. Dannenberg, W. P. Birmingham, B. Pardo, N. Hu, C. Meek, and G. Tzanetakis, "A comparative evaluation of search techniques for query-by-humming using the MUSART testbed," *Journal of the American Association for Information Science and Technology*, vol. 58, no. 5, pp. 687-701, 2007.

[69] J. Salamon, J. Serra, and E. Gomez, "Tonal representations for music retrieval: from version identification to query-by-humming," *International Journal of Multimedia Information Retrieval*, vol. 2, no. 1, pp. 45-58, 2013.

[70] J. Serra, E. Gomez, and P. Herrera, "Audio cover song identification and similarity: background, approaches, evaluation, and beyond," in *Advances in Music Information Retrieval*, Berlin, Germany: Springer, 2010, pp. 307-332.

[71] H. Kirchhoff and A. Lerch, "Evaluation of features for audio-to-audio alignment," *Journal of New Music Research*, vol. 40, no. 1, pp. 27-41, 2011.

[72] S. Dixon and G. Widmer, "Match: a music alignment tool chest," in *Proceedings of 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, 2005, pp. 492-497.

[73] M. Muller, H. Mattes, and F. Kurth, "An efficient multiscale approach to audio synchronization," in *Proceedings of 7th International Conference on Music Information Retrieval (ISMIR 2006)*, Victoria, Canada, 2006, pp.192-197.

[74] J. Lee, S. Park, S. Jo, and C. D. Yoo, "Music plagiarism detection system," in *Proceedings of the 26th International Technical Conference on Circuits/Systems, Computers and Communications*, Gyeongju, Korea, 2011, pp. 828-830.

[75] C. Dittmar, K. F. Hildebrand, D. Gaertner, M. Winges, F. Muller, and P. Aichroth, "Audio forensics meets music information retrieval: a toolbox for inspection of music plagiarism," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Bucharest, Romania, 2012, pp. 1249-1253.

[76] J. I. Park, S. W. Kim, M. Shin, "Music plagiarism detection using melody databases," in *Proceedings of the 9th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, Melbourne, Australia, 2005, pp. 684-693.

[77] A. Saracoglu, E. Esen, T. K. Ates, B. O. Acar, U. Zubari, E. C. Ozan, E. Ozalp, A. A. Alatan, and T. Ciloglu, "Content based copy detection with coarse audio-visual fingerprints," in *Proceedings of the 7th International Workshop on Content-Based Multimedia Indexing*, Chania, Greece, 2009, pp. 213-218.

[78] M. Schedl, E. Gomez, and J. Urbano, "Music information retrieval: recent developments and applications," *Foundations and Trends in Information Retrieval*, vol. 8, no. 2-3, pp. 127-261, 2014.

[79] R. Burke, "Hybrid recommender systems: survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331-370, 2002.

[80] Y. Song, S. Dixon, and M. Pearce, "A survey of music recommendation systems and future perspectives," in *Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval (CMMR 2012)*, London, UK, 2012, pp. 395-410.

[81] L. Baltrunas, M. Kaminskas, B. Ludwig, O. Moling, F. Ricci, A. Aydin, K. H. Luke, and R. Schwaiger, "InCarMusic: context-aware music recommendations in a car," in *Proceedings of the International Conference on Electronic Commerce and Web Technologies*, Toulouse, France, 2011, pp. 89-100.

[82] Y. C. Zhang, D. O. Seaghdha, D. Quercia, and T. Jambor, "Auralist: introducing serendipity into music recommendation," in *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, Seattle, WA, 2012, pp. 13-22.

[83] M. Schedl, D. Hauger, and D. Schnitzer, "A model for serendipitous music retrieval," in *Proceedings of the 2nd Workshop on Context-Awareness in Retrieval and Recommendation*, Lisbon, Portugal, 2012, pp. 10-13.

[84] M. Schedl and D. Schnitzer, "Hybrid retrieval approaches to geospatial music recommendation," in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Dublin, Ireland, 2013, pp. 793-796.

[85] D. Perrot and R. O. Gjerdigen, "Scanning the dial: an exploration of factors in the identification of musical style," in *Proceedings of the Society for Music Perception and Cognition*, Evanston, IL, 1999.

[86] F. Pachet and D. Cazaly, "A taxonomy of musical genres," in *Proceedings of RIAO 2000: Content-Based Multimedia Information Access*, Paris, France, 2000, pp. 1238-1245.

[87] G. Tzanetakis and P. Cook, "Marsyas: a framework for audio analysis," *Organised Sound*, vol. 4, no. 3, pp. 169-175, 2000.

[88] C. N. Silla, A. L. Koerich, and C. A. A. Kaestner, "A machine learning approach to automatic music genre classification," *Journal of the Brazilian Computer Society*, vol. 14, no. 3, pp. 7-18, 2008.

[89] T. Pohle, P. Knees, M. Schedl, E. Pampalk, and G. Widmer, "Reinventing the wheel: a novel approach to music player interfaces," *IEEE Transactions on Multimedia*, vol. 9, no. 3, pp. 567-575, 2007.

[90] D. Schnitzer, T. Pohle, P. Knees, and G. Widmer, "One-touch access to music on mobile devices," in *Proceedings of the 6th International Conference on Mobile and Ubiquitous Multimedia*, Oulu, Finland, 2007, pp. 103-109.

[91] N. Koenigstein and Y. Shavitt, "Song ranking based on piracy in peer-to-peer networks," in *Proceedings of 10th International Conference on Music Information Retrieval (ISMIR 2009)*, Kobe, Japan, 2009, pp. 633-638.

[92] M. Schedl, T. Pohle, N. Koenigstein, and P. Knees, "What's hot? Estimating country-specific artist popularity," in *Proceedings of 11th International Conference on Music Information Retrieval (ISMIR 2010)*, Utrecht, The Netherlands, 2010, pp. 117-122.

[93] R. Dhanaraj and B. Logan, "Automatic prediction of hit songs," in *Proceedings of 6th International Conference on Music Information Retrieval (ISMIR 2005)*, London, UK, 2005, pp. 488-491.

[94] I. Karydis, A. Gkiokas, and V. Katsouros, "Musical track popularity mining dataset," in *Proceedings of the 12th International Conference on Artificial Intelligence Applications and Innovations*, Thessaloniki, Greece, 2016, pp. 562-572.

[95] X. Serra, M. Magas, E. Benetos, M. Chudy, S. Dixon, A. Flexer, E., et al, *Roadmap for Music Information ReSearch*. London, UK: MIRES Consortium, 2013.

[96] Y. H. Yang and H. H. Chen, *Music Emotion Recognition*. Boca Raton, FL: CRC Press, 2011.

[97] P. Cano, E. Batlle, E. Gomez, L. de C. T. Gomes, and M. Bonnet, "Audio fingerprinting: concepts and applications," in *Computational Intelligence for Modelling and Prediction*. Berlin, Germany: Springer Science and Business Media, 2005, pp. 233-245.

[98] T. Kageyama, K. Mochizuki, and Y. Takashima, "Melody retrieval with humming," in *Proceedings of the International Computer Music Conference*, Tokyo, Japan, 1993, pp. 349-351.

[99] N. Kosugi, Y. Nishihara, T. Sakata, M. Yamamuro, and K. Kushima, "A practical query-by-humming system for a large music database," in *Proceedings of the 8th ACM international conference on Multimedi*, Los Angeles, CA, 2000, pp. 333-342.

[100] T. B. Mahieux and D. P. W. Ellis, "Large-scale cover song recognition using the 2D Fourier transform magnitude," in *Proceedings of 13th International Conference on Music Information Retrieval (ISMIR 2012)*, Porto, Portugal, 2012.

[101] M. P. Ryynanen and A. P. Klapuri, "Automatic transcription of melody, bass line, and chords in polyphonic music," *Computer Music Journal*, vol. 32, no. 3, pp. 72-86, 2008.

[102] K. Kashino, K. Nakadai, T. Kinoshita, and H. Tanaka, "Application of Bayesian probability network to music scene analysis," *Computational Auditory Scene Analysis*, vol. 1, no. 998, pp. 1-15, 1995.

[103] F. Pachet and P. Roy, "Hit song science is not yet a science," in *Proceedings of 9th International Conference on Music Information Retrieval (ISMIR 2008)*, Philadelphia, PA, 2008, pp. 355-360.

[104] J. S. Downie, K. West, A. Ehmann, and E. Vincent, "The 2005 Music Information Retrieval Evaluation eXchange (MIREX 2005): Preliminary Overview," in *Proceedings of 6th International Conference on Music Information Retrieval*, London, UK, 2005, pp. 320-323.

[105] M. D. Ferreira, D. C. Correa, M. A. Grivet, G. T. dos Santos, R. F. de Mello, and L. G. Nonato, "On accuracy and time processing evaluation of cover song identification systems," *Journal of New Music Research*, vol. 45, no. 4, pp. 333-342, 2016.

[106] R. Scholz, G. Ramalho, and G. Cabral, "Cross task study on MIREX recent results: an index for evolution measurement and some stagnation hypothesis," in *Proceedings of 17th International Conference on Music Information Retrieval (ISMIR 2016)*, New York, NY, 2016, pp. 372-378.

[107] K. Lemstrom and J. Tarhio, "Searching monophonic patterns within polyphonic sources," in *Proceedings of RIAO 2000: Content-Based Multimedia Information Access*, Paris, France, 2000, pp. 1261-1279.

[108] M. Grachten, J. L. Arcos, and R. L. de Mantaras, "Melodic similarity: looking for a good abstraction level," in *Proceedings of 5th International Conference on Music Information Retrieval (ISMIR 2004)*, Barcelona, Spain, 2004, pp. 210-215.

[109] N. Orio, "Music retrieval: a tutorial and review," *Foundations and Trends in Information Retrieval*, vol. 1, no. 1, pp. 1-90, 2006.

[110] P. Herrera-Boyer, X. Amatriain, E. Batlle, and X. Serra, "Towards instrument segmentation for music content description: a critical review of instrument classification techniques," in *Proceedings of 1st International Conference on Music Information Retrieval (ISMIR 2000)*, Plymouth, MA, 2000, pp. 115-119.

[111] P. Lamere, "Social tagging and music information retrieval," *Journal of New Music Research*, vol. 37, no. 2, pp. 101-114, 2008.

[112] A. Jha, "Music machine to predict tomorrow's hits," The Guardian, 2006 [Online]. Available: https://www.theguardian.com/technology/2006/jan/17/news.science.

[113] M. Goto, "Grand challenges in music information research," in *Multimodal Music Processing*, Dagstuhl, Germany: Dagstuhl Publishing, 2012, pp. 217-225.

Kuldeep Gurjar

He received B.S. (2005, Stany Mem. Collage) and M.S. degrees (2010) in Computer Science, from Department of Computer Science and Information Technology, University of Rajasthan, Jaipur. From 2010 to 2011, he worked for a Website development company (Octal info. Solutions). Since March 2012, he is with the Department of Computer Science and Engineering from Kangwon National University as a PhD candidate. His research interests are data mining, data provenance, data trustworthiness, music information retrieval, music copying detection.

Yang-Sae Moon https://orcid.org/0000-0002-2396-0405

He received B.S. (1991), M.S. (1993), and Ph.D. (2001) degrees in Computer Science from Korea Advanced Institute of Science and Technology (KAIST). From 1993 to 1997, he was a research engineer in Hyundai Syscomm, Inc., where he participated in developing 2G and 3G mobile communication systems. From 2002 to 2005, he was a technical director in Infravalley, Inc., where he participated in planning, designing, and developing CDMA and W-CDMA mobile network services and systems. He is currently a professor of computer science department at Kangwon National University. He was a visiting scholar at Purdue University in 2008 to 2009. His research interests include data mining, knowledge discovery, storage systems, access methods, multimedia information retrieval, big data analysis, mobile communication systems, and network communication systems. He is a member of the IEEE, and a member of the ACM.