*International Journal of Research in Advent Technology, Special Issue, August 2018*
*E-ISSN: 2321-9637*
*International Conference on "Topical Transcends in Science, Technology and Management" (ICTTSTM-2018)*
*Available online at www.ijrat.org*

# A Microscopic Comprehension of Contemporary Contribution on Music Retrieval System: An Analysis with Identical Tune

D. KHASIM VALI[1],
*Department of computer Science and Engineering, Vidyavardhaka College of Engineering, Mysuru – 570017, e-mail: dkv@vvce.ac.in.*
NAGAPPA U BHAJANTRI[2],
*Department of computer Science and Engineering, Government College of Engineering, Chamarajanagara – 571313, e-mail: bhajan3nu@gmail.com.*

**ABSTRACT**

The task of cover song identification is a relevant one in the field of Music Informatics Research. The music information retrieval (MIR) community has paid much attention to this task in recent years and many approaches have been proposed. In this work the comprehensively summarize the work done in Music information Retrieval and cover song identification with encompassing the background related to this area of research. The most promising strategies are reviewed and qualitatively comparedunder a common framework, and their evaluation methodologies are critically addressed.

**Keywords:** identical tune, MIR, MIREX, FFT

## 1. INTRODUCTION

Music is now a digital resource hugely available in various media like mobile phones, digital audio players, pen-drives, websites etc. The volume of music available makes it difficult for a listener to select a desired piece of music. Advanced techniques need to be developed in order to facilitate efficient searching. Traditional methods of searching for a preferred music track by singer, lyrics or by composer are to be replaced by more deeper search mechanisms where a listener can search for a piece of music by humming a part of the tune. This is called 'Query By Humming (QBH)' [1], an area of research coming under data mining of music. There exists immense scope for research in the area of discovering trends and patterns in music. Methods for more sophisticated access to music collections need to be developed in order to enhance search and browse functionality. The area of research for developing such methods is called Music Information Retrieval (MIR). It is an area of intensive research by academic and industrial research laboratories, archives, and libraries [2].

The focus of research in Computational Musicology(CM) and MIR is not to study music as such, but to design methods to retrieve music from large databases using its musical content rather than metadata. Tasks involved in CM and MIR include genre classification, raga recognition, melody extraction, artist recognition, song recommendation etc. The benefits of MIR reach out to various categories like recording industry, listeners searching for preferred music, professionals including music performers, teachers, musicologists, copyright lawyers, music producers etc [2].

### 1.1 Metadata-Driven MIR

The present MIR systems are mainly metadata-driven. In other wards, data is about represents largely metadata. That is, textual information such as lyrics, artist, composer, album, year etc [3]. There are many music searching services using metadata only. For example, *pandora.com* is a commercial metadata-driven music system in which the user can search for a particular song by specifying the artist name or song name and a radio station, featuring that song and songs similar to it, will be created by the system. The system works on metadata entered externally for each of the tracks. However, when catalogues become very large it becomes extremely difficult to maintain the metadata descriptions. Also, editorial supervision of the metadata is necessary [2].

The success of a metadata based MIR system depends on the accuracy of the descriptions about music contained in its repository. Incorrect data on artist, album, year, track title, lyrics, duration etc, can severely affect the utility of metadata based MIR systems. Due to these limitations, metadata based MIR systems have only limited usability. Even commercial MIR systems using metadata are not able to provide users with search capabilities for finding music which they do not know how to search for.

The limitations of centralized metadata can be overcome, to some extent, by allowing the users to update the metadata content. For example, websites like *gracenote dotcom* and *musicbrainz dotorg* provide metadata contributed by user communities. Similarly, users of websites like *myspace dotcom*, *flickr dotcom*, and *youtube dotcom* can find items of interest indexed by users with similar tastes.

### 1.2 Content-Based MIR

The limitations of metadata based MIR systems lead to the research on content-based methods which enable users to access music using content based search. Content-based MIR systems enable a user to seek a desired piece of music even when user does not know exactly what the interest is. For example, there are occasions when user know the tune of a song but cannot remember its lyrics, artist, album, year, composer or any information that is necessary as input in a metadata based MIR system.

In content based MIR systems, a user can give a tune hummed by him as input and get the desired song with that tune as output. For example, the website *shazam dotcom* described in [4], can identify a particular recording from an input sample and output the artist, album, and track title. Similarly in online music service *nayio dotcom*, a user can sing a query to identify the musical piece.

The main tasks in content based MIR are query formation, description extraction, matching and, finally, music document retrieval. Use cases define the type of query, the sense of match, and the form of the output. Queries and output can be music fragments, recordings, scores, or music features. The degree or accuracy of the match is called specificity of the MIR system. Specificity is exact when music with specific content is retrieved and approximate when music along with near neighbors is retrieved. Content based MIR systems can be divided into three categories based on their specificity. Systems that identify exact content are called high-specificity systems.

High-specificity systems match instances of audio signal content. That is, they perform an exact content level comparison between the audio signal in the input query and the audio signal in the database. Systems that return music with some matching global characteristics are called low specificity systems. They make use of broad descriptions of music, such as genre. The tracks returned by such systems will have only a broad relationship in common with the query. The third type of content based MIR systems is mid-specificity systems which match high-level music features, such as melody, but do not match audio content.High-specificity system includes plagiarism detection, copyright monitoring etc. Mid-specificity system is artist identification, composer identification, melody identification, raga recognition and classification etc. Examples of low-specificity system are emotion identification, genre detection, instrument detection etc.

## 2. REPRESENTATION OF MUSIC

Representation of music is an important factor in MIR. There are two types of musical representations in content-based MIR: audio data and symbolic data. Audio data means a sampled and quantized sound wave whereas symbolic data means some kind of higher level representation of music in which notes, pitches, durations, instruments etc are explicitly encoded. An audio representation is closer to the sound that reaches the ear, while a symbolic representation is closer to notated music [3]. For studies involving musical scores, symbolic representation is ideal, since it is the most direct representation of the material that is studied. Audio data is ideal for content based search and other information retrieval activities.

Tasks involved in audio based MIR are beat detection, tempo tracking, pitch detection, source separation such as separating the violin from the vocal etc. Extracting musical information from the content of an audio signal is a very difficult task. Even in the case of a single instrument, or an unaccompanied singer, detection of basic properties like the pitch that sounds at a certain moment in time is very difficult [3]. For more complex signals representing the sound of many instruments sounded together, the problem becomes more complicated. Although it is difficult to obtain musical information from audio, such kind of extracted information will be more similar to the actual musical feeling experienced by human ear compared to symbolic information.The musical content information handled by content based MIR systems can be of two types: High-level music content and Low-level audio features.

### 2.1 High-Level Music Content Information

High-level musical information includes musical concepts such as melody, rhythm, timbre and harmony. They are used to describe the content of the music. Melody enables us to distinguish one work from another and to reproduce it by singing, humming or whistling. Melody makes music memorable. That is how we recall a tune long after we have forgotten its text [5]. Query-by-humming systems try to extract melodic content from audio signals so that a user can search for music by singing or humming part of the melody. As said earlier, an example for such systems is *naiyo dotcom*. A survey of sung-query methods conducted by Hu and Dannenberg is described in [6].

Extraction of high-level music concepts such as melody, rhythm, timbre and harmony is a sub-goal of MIR and is a subject of intensive research. It is an extremely difficult task especially from polyphonic recordings, i.e., multiple instruments playing simultaneously.

***2.2 Low-Level Audio Features***

The existing symbolic standards such as Musical Instrument Digital Interface (MIDI) which provides symbolic musical information are not comprehensive enough to represent the nuances of non-western music especially Indian classical music. MIDI cannot reproduce the ornamentations called '*gamakas*' which are essential to get the *'raga bhavam'* (true color of a raga) in most of the Indian ragas. To build a computational model that can comprehensively represent Indian classical music, we need to do deeper, low-level analysis on the audio recording itself. Also, there are a great number of Indian classical music recordings available that can be used as input. This motivates work on extracting high-level music features from low-level audio content. That is, extracting information contained in the digital audio.

The Music Information Retrieval Experimental Exchange (MIREX) [7] community who studies automatic extraction of factual, cultural, and high-level music descriptions has found that, even in the case of Western music, low-level audio methods outperform symbolic methods even when clean symbolic information such as MIDI is available.

Low-level audio features are quantitative measures of audio signals such as fundamental frequency, amplitude etc that containencoded musical information. Generally, methods for extraction of low-level audio features can be categorized into three types:

a) Frame based methods where the input audio signal is segmented into frames with duration of 10 ms-1000 ms intervals.
b) Beat-synchronous methods where features are aligned to musical beat boundaries.
c) Statistical measures where probability distributions using the extracted features are constructed for analysis.

***2.2.1 Techniques for the Extraction of Low-Level Audio Features***

Various techniques employed to extract low-level audio features from digital audio music signal are Autocorrelation, Fast Fourier Transform (FFT), Short-Time Magnitude Spectrum, Constant-Q/Mel Spectrum, Pitch-Class Profile, Onset Detection, Mel/Log-Frequency Cepstral Coefficients (MFCC) and Spectral Flux.

**i) Fast Fourier Transform (FFT)**

The Fast Fourier Transform (FFT) is a powerful tool widely usedin signal processing. In musical analysis, it is used for extracting the spectral information of a signal, such as in pitch tracking. The FFT can be combined with the Inverse Fast Fourier Transform (IFFT). The application of the FFT/IFFT in MIR offers a high degree of control on a given signal's spectral information making it possible for flexible and efficient implementation of MIR algorithms.

**ii) Autocorrelation**

Autocorrelation is a mathematical tool for finding repeatingpatterns. Informally, it is the similarity between observations as afunction of the time separation between them. It is often used in signal processing for analyzing functions or series of values, such as time domain signals. In musical analysis, it is an effective tool for applications such as identifying the fundamental frequency in a musical signal implied by its harmonic frequencies and separating a periodic signal from noise etc.

**iii) Short-Time Magnitude Spectrum**

Low-level audio features can be extracted based on a short-time spectrum of the audio signal. Since the phase is not as perceptually salient for music as the magnitude, the magnitude spectrum can be used for feature extraction.

**iv) Constant-Q/Mel Spectrum**

When sound enters the ear, it causes vibrations on the basilarmembrane within the inner ear. Different frequencies of sound cause different regions of the basilar membrane and its fine hairs to vibrate. This is how the brain discriminates between various frequencies. However, if two frequencies are close together, there is an overlap of response on the basilar membrane and these frequencies cannot be distinguished as separate frequencies. Instead an average frequency isheard. Here, the two frequencies are said to be within a critical band on the basilar membrane. Critical bands are non-uniform frequency bands that group close frequencies into a single band of a given center frequency. These constant bandwidth critical bands can be represented using a constant-Q transform, where the Q is the ratio of bandwidth to frequency [6]. Since musical notes occur in octaves, a fraction of the octave, for example twelfth (one semitone in Western music), can be used for band centers. The Mel or Constant-Q spectrum can be obtained from a linear spectrum by summing the powers in adjacent frequency bands. This approach employs FFT to compute the spectrum.

**v) Pitch-Class Profile**

There are 12 equally spaced (equal temperament) pitch classes inWestern tonal music. Pitch-class profile (PCP) folds the frequency corresponding to each pitch into one of these 12 pitch classes [17]. So there are typically 12 bands in a PCP representation. This method integrates the energy in all octaves corresponding to one pitch class into a single band, independent of pitch height. If the octave is divided into an integer multiple of 12 such as 24, 36, or 48 bands, we can increase resolution of pitch information.

The PCP method employing equally spaced pitch classes, which works well for western music, are not ideal as such in case of non-western music. Extraction of features by using unequally spaced pitch classes is necessary for application to non-western music such as Carnatic music. Also, ideal numbers of pitch classes for Carnatic music are 16 and 22. For the recognition of 72 *melakartha* ragas and some *janya*ragas in Carnatic Music, 16 classes are needed. To identify someother *janya* ragas like *Saveri*, *Begada*etc, more than 16 pitch classes are needed. Theoretically, there are 22 *sruthis* in Carnatic music and hence a complete raga recognition system will have to incorporate these 22 *sruthis*.

**vi) Onset Detection**

A Musical note has an attack (beginning) followed by sustain(duration) and decay (end) portions. Onset denotes the point where a musical note starts. Onset detection is concerned with marking the beginnings of notes, that is, to find out the boundaries between notes. There are several methods for onset detection. Some use the spectral differences in the magnitude spectrum of adjacent time points, others use phase differences in adjacent time points. Combination of both spectral difference and phase difference is also used and it is called complex number onset detection [16].

**vii) Mel/Log-Frequency Cepstral Coefficients (MFCC)**

This is a real-valued implementation of the complex Cepstrum insignal processing [32]. In this method, the logarithm of the Mel magnitude spectrum is taken and the resulting values are de-correlated using a Discrete Cosine Transform. The result is a real-valued array containing sinusoidal modulation of spectral magnitudes in the increasing order of modulation frequency.

**viii) Spectral Flux**

The rate of change of the power spectrum of a signal is called the spectral flux. It is estimated by comparing the power spectrum of a signal frame with the power spectrum of a previous frame. The spectralflux of a musical signal estimates the fine spectral-temporal structure in different frequency bands by measuring the modulation amplitudes in mid-to-high spectral bands [8]. The resultant feature vector is a two-dimensional matrix, with frequency bands in the rows and modulation frequency in columns, representing the rate of change of power in each spectral band.

#### *2.2.2 Aggregate Representations*

After extracting the low-level audio features, the next step in MIR research is to convert these low-level features into aggregate representations. Aggregate features are needed for obtaining high-level musical information such as similarity measurement, raga recognition, tempo tracking, beat detection etc. Extraction of the low-level audio features and their aggregate representations are the first two steps in music information retrieval where the ultimate aim is to obtain a high-level representation of music for the processing of music content.

## 3. PROBLEM PATTERN

However on above information on MIR system. Here, community effort on MIR has explosed in maticulrly aremotivated to doresearch work predominantly establishes interest to expand initiative to extract tune similarity of music information from the wider categories of music database.Here we review cover song identification system.

### 3.1 Cover Song Identification

Cover song identification has been a very active area of study within the last few years in the MIR community, and its relevance can be seen from multiple points of view. From the perspective of audio content processing, cover song identification yields important information on how musical similarity can be measured and modeled. Music similarity is an ambiguous term and, apart from musical facetsthemselves, may also depend on different cultural (or contextual) and personal (or subjective) aspects [9]. The purpose of many studies is to define and evaluate the concept of music similarity, but there are many factors involved in this problem, and some of them (maybe the most relevant ones) are difficult to measure [10].

Still, the relationship between cover songs is context-independent and can be qualitatively defined and objectively measured, as a "canonical" version exists and any other rendition of it can be compared to that. The problem of identifying covers is also challenging from the point of view of music cognition, but apparently it has not attracted much attention by itself. When humans are detecting a cover, they have to derive some invariant representation of the whole song or maybe of some of its critical sections.

From a commercial perspective, it is clear that detecting cover songs has a direct implication to musical rights' management and licenses. Furthermore, quantifying music similarity is key to searching, retrieving, and organizing music collections. Nowadays, online digital music collections are in the order of ten [11] to a few hundred million tracks and they are continuously increasing. Therefore, one can hypothesize that the ability to manage this huge amount of digital information in an efficient and reliable way will make the difference in tomorrow's music-related industry [12, 13]. Personal music collections, which by now can easily exceed the practical limits on the time to listen to them, might benefit as well from efficient and reliable search and retrieval engines. From a user's perspective, finding all versions of a particular song can be valuable and fun. One can state an increasing interest for cover songs just by looking at the emergence of related websites, databases, and podcasts in the internet such asSecond Hand Songs3 , Coverinfo4 , Coverville5 , Midomi6 , Fancovers7 , or YouTube8 . Frequently, these sites also allow users to share/present their own (sometimes homemade) cover songs, exchange opinions, discover new music, make friends, learn about music by comparing versions, etc. Thus, cover songs are becoming part of a worldwide social phenomena.

**3.2 Approaches**

The standard approach to measuring similarity between cover songs is essentiallyto exploit music facets shared between them. Since several important characteristicsare subject to variation (timbre, key, harmonization, tempo, timing, structure, and soforth), cover song identification systems must be robust against thesevariations.

Extracted descriptors are often in charge of overcoming the majority of musicalchanges among covers, but special emphasis is put on achieving tempo, key, or structureinvariance, as these are very frequent changes that are not usually managed byextracted descriptors themselves. Therefore, one can group the elements of existingcover song identification systems into four basic functional blocks: feature extraction,key invariance, tempo invariance, and structure invariance. An extra block canbe considered at the end of the chain for the final similarity measure used (figure 1illustrates these blocks).

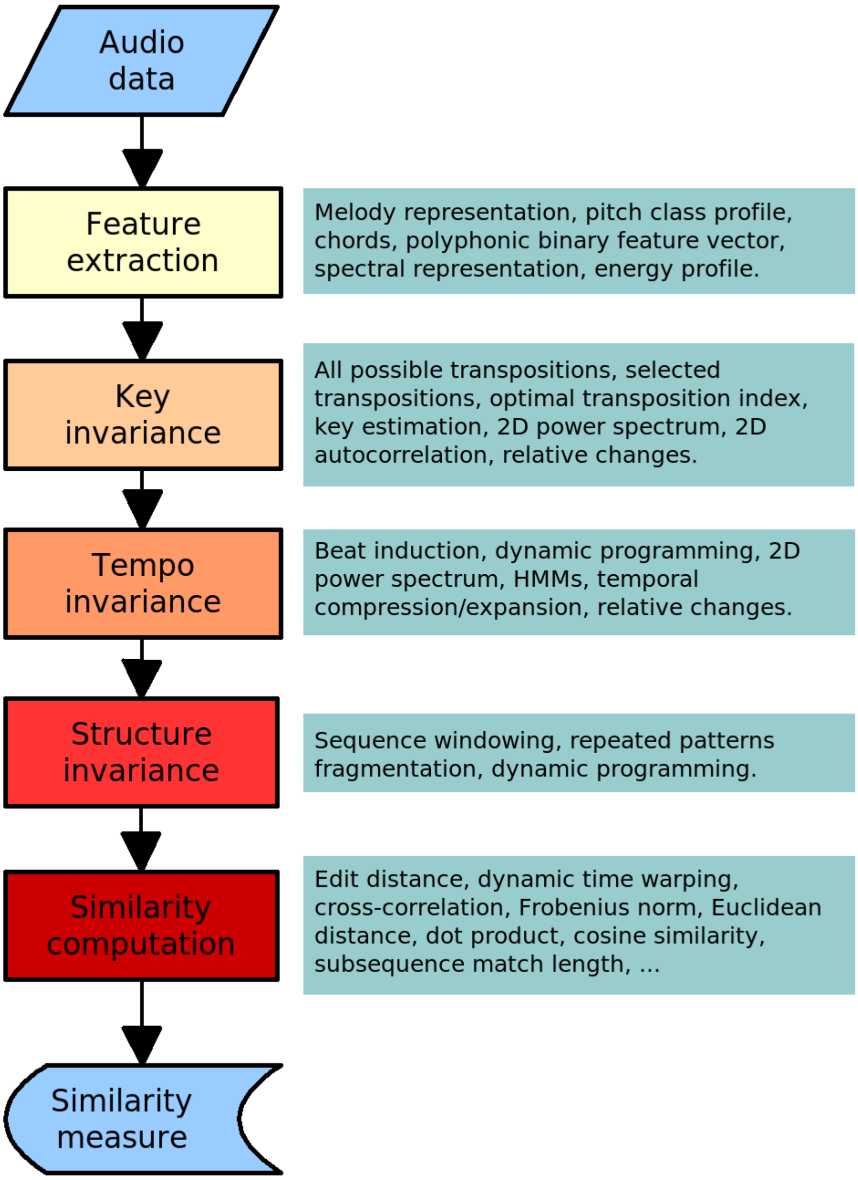


**Fig. 1 Generic block diagram for cover song identification systems.**

## 4. EVALUATION

The evaluation of cover song identification and similarity systems is a complex task,and it is difficult to find in the literature a common methodology for that. The onlyexisting attempt to compare version identification systems is found in the MusicInformation Retrieval Evaluation eXchange (MIREX) initiative. Nevertheless,the MIREX framework only provides an overall accuracy of each system. Avaluable improvement would be to implement independent evaluations for the differentprocesses involved (feature extraction, similarity computation, etc.), in orderto analyze their contributions to the global system behavior.

The evaluation of cover song identification systems is usually set up as a typicalinformation retrieval "query and answer" task [14], where one submits a querysong and the system returns a ranked set (or list) of answers retrieved from a givencollection [15]. Then, the main purpose of the evaluation process is to assess howprecise the retrieved set is. We discuss two important issues regarding the evaluation of cover song retrieval systems: the evaluation measures and the music materialused.

## 5. CONCLUSION

We have summarized here the work done for addressing the problem of Music information retrieval. In addition, we also discussed the motivation to do research work automaticcover song identification. Even though different approaches have been tried, it seemsquite clear that a well-crafted system has to be able to exploit tonal, temporal, andstructural invariant representations of music. We have also learnt that there aremethodological issues to be considered when building music collections used asground-truth for developing and evaluating cover identification systems.

**Table 1: Shows the overall work in Music Information Retrieval**

| Sl. No | Author and year | Method used | Features | Complexity | Metric performance inability | Short comes and remarks |
|---|---|---|---|---|---|---|
| 1 | Mitali Bagul, Divya Soni, Saravanakumar K-2014 | Common melody extraction by tempo/BPM and Spectrogram | Tempo/BPM (Beats Per Minute) | 1. Songs has max frequency called Pleasant Limit (PL).<br>2. Pleasant Limit makes the song popular. | Result shows that no song in the dataset exceeds this limit. | Language based Identification |
| 2 | Tzu-Hsiang Huang, pao-Chi Chang-2016 | Machine learning approaches such as pooling and Cosine distance similarity | Chroma features, loudness, beat information | 1. K-means algorithm is used.<br>2. PCA (Principal Component Analysis) and LDA (Linear Discriminate Analysis) are used. | 2D Fourier transform used to compress the music information. | Less accuracy |
| 3 | Kang Cai, Deshun Yang, Xiaoou Chen-2016 | Two-layer approach, first layer is Fast rejector and the second layer is Segment Cross Similarity | 2D-FMC over a patch with fixed l beats | 1. Fast rejector rejects non-coves efficiently.<br>2. Structural Cross Similarity has three stages Such as Music segmentation, Section feature extraction and similarity measurement. | Proposed method significantly improves the mean AP of the dataset and also improves the performance. | Neglects the inner relationship among the sections of the song. |
| 4 | Antonio CAmarena-Ibarrola, Karina Figueroa, Hector Tejada-Villela-2016 | Sensitivity Analysis | sequence of entropy-per chroma vectors | 1. Threshold is used to compare the distance between two degraded versions of the same song. | Entropy Chromagrams in cover song identification is efficient. | Less accurate for larger databases. |
| 5 | Swathi Chauhan, Prachi Chauhan-2016 | Latent Dirichlet Allocation(LDA) | Lyrics | 1. Classified based on Class-H and Class-S mood.<br>2. LDA has four stages such as Document-Term | Performs lyrics analysis to recognize the mood-sentiment from hindi lyrics using | No existing dataset or result to evaluate the model output. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | Matrix, Fitting the model, Topic-word distribution and Document-Topic distribution. | unsupervised machine learning approach. | |
| 6 | Peter Foster, Simon Dixon, Anssi Klapuri-2015 | Information-theoretic approach and Normalized Compression Distance with Alignment(NCDA) | 12-component chroma features, alongside predicted note and beat onsets. | 1. NCDA incorporates a method of obtaining joint representation of time series based on cross-prediction. | Efficiently output the evaluated measure of pair wise predictability between time series for cover song identification. | Fail to evaluate alternative time series. |
| 7 | Maksim Khadkevich, Maurizio Omologo-2013 | Locality Sensitive Hashing(LHS) | Chroma features. | 1. Chord progressions and chord profiles are used as high-level descriptors. | Scalable in identifying the cover songs. | Low Performance |
| 8 | Marko Stamenovic-2015 | Automatic extraction of audio features with a stacked auto-encoder (SAE) combined with beat tracking. | Tempo | 1. Distance measurement is done between SAE feature and ground truth. | Automatically classifies cover songs based on spectral analysis and based on SAE. | Fails to identify highly dissimilar songs. |
| 9 | Eric J.Humphrey, Oriol Nieto, Juan P.Bello-2013 | Feed-forward architectural approach | Beat-chroma patches. | 1. Uses data-driven projections at different time scales to compute local features | Uses 2D-FMC method and also applies LDA to baseline system | Over-fitting problem in higher dimensions. |
| 10 | J.Osmalskyj, S.pierard, M. Van Droogenbroeck, J.J. Embrechts-2013 | mixes independent rejectors together to build Composite ones. | Tempo and Chroma Features. | 1. Elementary Rejectors are Duration and tempo rejector, chroma rejevtor. | To evaluate the performances of a rejector, one has to consider both the pruning rate and the risk to reject all the corresponding versions | Overall performance is less. |

**REFERENCES**

[1] J. Li J. Han Z. Shi J. Li "An Efficient Approach to Humming Transcription for Query-by-Humming System", 3rd International Congress on Image and Signal Processing, 2010.
[2] Casey, M.A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., Slaney, M., Content-Based Music Information Retrieval: Current Directions and Future Challenges, Proceedings of the IEEE, Vol. 96(4), 2008, pp. 668-696.

[3] Kranenburg, P.V., A Computational Approach to Content-Based Retrieval of Folk Song Melodies, Ph.D Thesis, Utrecht University, Netherlands.
[4] Wang, A., The Shazam Music Recognition Service, Communications of the ACM-Music Information Retrieval, Vol. 49(8), 2006, pp. 44-48.
[5] Hewlett, W.B., Selfridge-Field, E., Conceptual and Representational Issues in Melodic Comparison, Melodic Similarity - Concepts, Procedures, and Applications, MIT Press, 1998.
[6] Hu, N., Dannenberg, R., A Comparison of Melodic Database Retrieval Techniques Using Sung Queries, Joint Conference on Digital Libraries, New York, ACM Press, 2002, pp. 301-307.
[7] Salamon, J., Gulati, S., Serra, X., A Multipitch Approach to Tonic Identification in Indian Classical Music, 13th International Society for Music Information Retrieval Conference (ISMIR 2012), Porto.
[8] Pampalk, E., Computational Models of Music Similarity and Their Application in Music Information Retrieval, Ph.D. Dissertation, Vienna University of Technology, Vienna, Austria, 2006.
[9] D. P. W. Ellis, B. Whitman, A. Berenzweig, and S. Lawrence. The quest for ground truth in musical artist similarity. Int. Symp. on Music Information Retrieval (ISMIR), pages 518–529, October 2002.
[10] A. Berenzweig, B. Logan, D. P. W. Ellis, and B. Whitman. A large scale evaluation of acoustic and subjective music similarity measures. Int. Symp. on Music Information Retrieval (ISMIR), 2003.
[11] F. Pachet. Knowledge management and musical metadata. Idea Group, 2005.
[12] M. Casey, R. C. Veltkamp, M. Goto, M. Leman, C. Rhodes, and M. Slaney. Content-based music information retrieval: current directions and future challenges. Proceedings of the IEEE, 96(4):668–696, April 2008
[13] F Vignoli and S. Paws. A music retrieval system based on user-driven similarity and its evaluation. Int. Symp. on Music Information Retrieval (ISMIR), pages 272–279, 2005.
[14] R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. ACM Press Books, 1999.
[15] J. S. Downie, M. Bay, A. F. Ehmann, and M. C. Jones. Audio cover song identification: Mirex 2006-2007 results and analyses. Int. Symp. on Music Information Retrieval (ISMIR), pages 468–473, September 2008
[16] T. Dharani, and I. L. Aroquiaraj, "A survey on content based image retrieval," in International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME 2013) Salem, USA, 2013, pp. 485-490.
[17] Y. Gong, S. Lazebnik, A. Gordo, and F. Perronnin, "Iterative quantization: a procrustean approach to learning binary codes for large-scale image retrieval," IEEE Transaction on Pattern Analysis and Machine Intelligence vol. 35, no. 12, pp. 2916-2929, 2013.
[18] J. S. Downie. "The International Society of Music Information Retrieval," http://www.ismir.net/.
[19] Z. W. Ras, and A. Wieczorkowska, Advances in Music Information Retrieval, 1 ed.: Springer-Verlag Berlin Heidelberg, 2010.
[20] M. Schedl, E. Gómez, and J. Urbano, "Music information retrieval: recent developments and applications," Foundations and Trends in Information Retrieval, vol. 8, no. 3, pp. 127-261, 2014.
[21] 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.
[22] M. Helén, and T. Virtanen, "Audio query by example using similarity measures between probability density functions of features," EURASIP Journal on Audio, Speech, and Music Processing, pp. 1-12, 2010.
[23] W.-H. Tsai, H.-M. Yu, and H.-M. Wang, "Query-by-example technique for retrieving cover versions of popular songs with similar melodies," in 6th International Conference on Music Information Retrieval, London, UK. September 11-15, 2005, pp. 183-190.
[24] I. S. H. Suyoto, A. L. Uitdenbogerd, and F. Scholer, "Effective retrieval of polyphonic audio with polyphonic symbolic queries," in MIR '07 Proceedings of the International Workshop on Multimedia Information Retrieval, 2007, pp. 105-114.
[25] J. Makhoul, F. Kubala, T. Leek, D. Liu, L. Nguyen, R. Schwartz, and A. Srivastava, "Speech and language technologies for audio indexing and retrieval," Proceedings of the IEEE, vol. 88, no. 8, pp. 1338-1353, 2000.
[26] W.-H. Tsai, Y.-M. Tu, and C.-H. Ma, "An fft-based fast melody comparison method for query-by-singing/humming systems," Pattern Recognition Letters, vol. 33, pp. 2285-2291, 2012.
[27] H.-M. Yu, W.-H. Tsai, and H.-M. Wang, "A query-by-singing system for retrieving karaoke music," IEEE Transactions on Multimedia vol. 10, no. 8, pp. 1626-1637, 2008.

[28] E. Unal, E. Chew, P. G. Georgiou, and S. S. Narayanan, "Challenging uncertainty in query by humming systems: a fingerprinting approach," IEEE Transactions on Audio, Speech and Language Processing, vol. 16, no. 2, pp. 359-371, 2008.
[29] M. Helén, and T. Virtanen, "A similarity measure for audio query by example based on perceptual coding and compression," in Proceedings of the 10th International Conference on Digital Audio Effects (DAFx-2007), Bordeaux, France, September 10-15, 2007, pp. 1-4.
[30] [30] K. Itoyama, M. Goto, K. Komatani, and T. Ogata, "Query-by-example music information retrieval by score informed source separation and remixing technologies," EURASIP Journal on Advances in Signal Processing, pp. 1-14, 2010.
[31] Y. Vaizman, B. McFee, and G. Lanckriet, "Codebook-based audio feature representation for music information retrieval," IEEE/ACM Transactions on Audio, Speech and Language Processing, vol. 22, no. 10, pp. 1483-1493, 2014.
[32] J. Salamon, "Pitch analysis for active music discovery," in 33rd International Conference on Machine Learning, New York, 2016, pp. 1-3.
[33] B. Thoshkahna, and K. Ramakrishnan, "Arminion:a query by example system for audio retrieval," Proceedings of Computer Music Modelling and Retrieval, pp. 1-9, 2005.
[34] A. Schröder, and M. Keith. "Free database," http://www.freedb.org.
[35] R. Kaye. "the Open Music Encyclopedia," https://musicbrainz.org.
[36] C. Barton, P. Inghelbrecht, A. Wang, and D. Mukherjee. "Shazam Company," http://www.shazam.com/company.
[37] F. Chuffart. "Musiwave "; http://www.musiwave.net.
[38] J. Born. "Neuros "; www.neurostechnology.com.
[39] K. Mohajer, M. Emami, J. Hom, K. McMahon, T. Stonehocker, C. Lucanegro, K. Mohajer, A. Arbabi, and F. Shakeri. www.soundhound.com.
[40] M. Gowan. http://www.techhive.com/
[41] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, Digital Watermarking and Steganography, 2 ed.: Morgan Kaufmann, 2007.
[42] H. Harb, and L. Chen, "A query by example music retrieval algorithm," in in Proceedings of the 4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '03), 2003, pp. 1-7.
[43] M. Bagul and D. Soni and K. Saravanakumar," Recognition of similar patterns in popular Hindi Jazz songs by music data mining", International Conference on Contemporary Computing and Informatics (IC3I), pp 1274-1278, 2014
[44] T. H. Huang and P. C. Chang, Large-Scale Cover Song Retrieval System Developed Using Machine Learning Approaches IEEE International Symposium on Multimedia (ISM), pp 591-596, 2016.
[45] K. Cai and D. Yang and X. Chen, Two-layer large-scale cover song identification system based on music structuresegmentation, IEEE 18th International Workshop on Multimedia Signal Processing (MMSP),pp 1-6,2016.
[46] A. Camarena-Ibarrola and K. Figueroa and H. Tejeda-Villela, "Entropy per chroma for Cover song identification", IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC), pp 1-6,2016.
[47] S. Chauhan and P. Chauhan, Music mood classification based on lyrical analysis of Hindi songs using Latent Dirichlet Allocation, International Conference on Information Technology (InCITe) - The Next Generation IT Summit on the Theme - Internet of Things: Connect your Worlds, pp 72-76, 2016.
[48] P. Foster and S. Dixon and A. Klapuri, "Identifying Cover Songs Using Information-Theoretic Measures of Similarity", IEEE/ACM Transactions on Audio, Speech, and Language Processing, pp 993-1005 , 2015
[49] Maksim Khadkevich and Maurizio Omologo, Large-Scale Cover Song Identification Using Chord Profiles, International Society for Music Information Retrieval, 2013
[50] Marko Stamenovic, Identifying Cover Songs Using Deep Neural Networks, International Symposium on Multimedia (ISM), pp 51-56, 2014.
[51] Eric J Humphrey, Oriol Nieto, Juan P Bello Data Driven and Discriminative Projections for Large-scale Cover Song Identification*Proc. of the 14th International Society for Music Information Retrieval Conference* Curitiba, Brazil,2013

[52] Osmalskyj and S. Piérard and M. Van Droogenbroeck and J. J. Embrechts, Efficient database pruning for large-scale cover song recognition, IEEE International Conference on Acoustics, Speech and Signal Processing, pp 714-718, 2013