ISSN 2278-3091
Volume 9 No.2, March -April 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse11922020.pdf
https://doi.org/10.30534/ijatcse/2020/11922020

A Review: Music Feature Extraction from an Audio Signal

Mukkamala S N V Jitendra[1], **Dr Y. Radhika**[2]
[1, 2]Department of Computer Science and Engineering,
GIT, GITAM University,
Visakhapatnam, AP, India.
jitendra.mukkamala@gmail.com, radhika.yalavarthi@gitam.edu

ABSTRACT

In recent years, the revenue earned through digital music stood at a billion-dollar market and the US remained the most profitable market for Digital music. Due to the digital shift, today people have access to millions of music clips from online music applications through their smart phones. In this context, there are some issues identified between the music listeners, music search engine by querying and retrieving music clips from a large collection of music data set. Classification is one of the fundamental problems in music information retrieval (MIR). Still, there are some hurdles according to their listener's preferences regarding music collections and their categorization. In this paper, different music extraction features are addressed, which can be used in various tasks related to music classification like a listener's mood, instrument recognition, artist identification, genre, query-by-humming, and music annotation. This review illustrates various features that can be used for addressing the research challenges posed by music mining.

Key words: Music mining, music information retrieval, feature extraction, listener's mood, instrument recognition, artist identification, genre, query-by-humming, and music annotation.

1. INTRODUCTION

Music always plays a major role in human entertainment. With the advent of digital music and cloud technologies, the vast amount of music accessible through online streaming, for this intent it is necessary to develop tools to improve the music navigation, efficient retrieval and to manage the music clips according to the user's interest. Furthermore, huge amounts of available music data have opened new opportunities for researchers working on music information retrieval and recommendation systems. It is high due to the economic potential of online music content.

The following statistics back the dominance of digital music; Global music report 2019 says the revenue earned through digital music stood at $19.1 billion in 2018, Global digital music revenue through downloads in 2019 - the US $2,065 million and Global digital music revenue through streaming in 2019 - $11,110 million. Revenue is expected to show an annual growth rate (CAGR 2019-2023) of 4.1%, resulting in a market volume of US $13,051 million by 2023. User penetration is 14.0% in 2019 and is expected to hit 15.3% by 2023. Fig.1. Shows the statistics of global recorded music industry revenues in between 2001-2018. Open-source music apps such as Spotify, Pandora, Shazam, Sound Cloud, Amazon Music, Google Play Music, Wynk Music,etc. Have been a great hit around the globe.There are various approaches like content-based information retrieval which helps in retrieving textual data while in view of querying multimedia like images, audio and video content-based multimedia information retrieval is used.

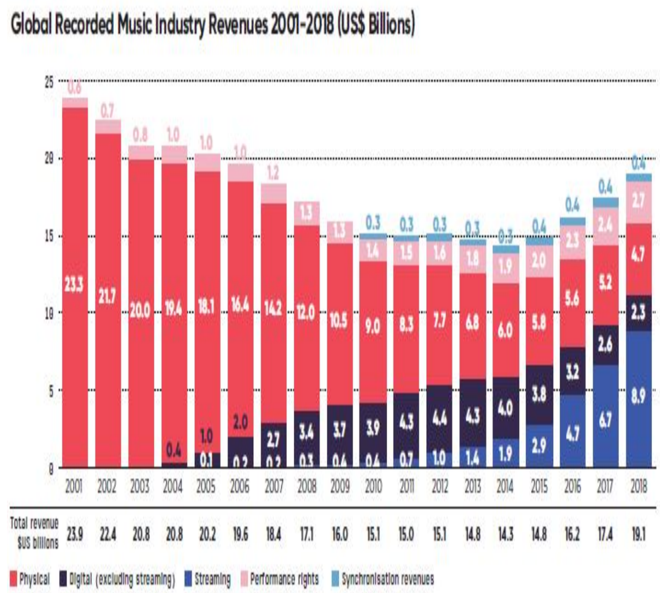

Figure 1: Growth of global recorded music industry revenue.

Source:https://www.ifpi.org/news/IFPI-GLOBAL-MUSIC-REPORT-2019.

Similarly, music mining is an interesting topic with many potential applications. Music mining addresses the field of data mining which helps in retrieving useful information from outsized collections of music. It can be viewed as a compartment of music information retrieval (MIR) and it can

be referred to as content-based music information retrieval (CB-MIR).
In the noticing years, the field of music information retrieval grew to cover a wide range of techniques for music analysis. For computers, unlike humans, music is nothing else than a form of an audio signal. Therefore, MIR uses audio signal analysis to extract meaningful features of music and to classify according to the listener's preferences.

2. Audio Features And Classifiers For Music Mining

The backbone of most music information retrieval systems is the features extracted from audio [42]. High quality audio songs are mainly available online with a sampling frequency in the range between 96kHz -192 kHz (24-bit) and for compact disc, it us 44.1kHz (16-bt). Direct processing of these high-quality audio songs for information retrieval utilizes large memory and processing time. Mostly, numeric features are extracted that resemble the signal characteristics and succinctly represent the original audio songs. There are a tremendous number of features that have been presented with the support of various signal processing techniques and statistical methods to simplify the tasks of speech processing. The majority of them are used to characterize the music easily. Some additional features are also introduced to model the music signal in a better manner. The principal components of a classification system are feature extraction and classifier learning. Feature extraction addresses the problem of how to represent the instances to be classified in terms of feature vectors or pair wise similarities. The purpose of classifier learning is to find a representative of the feature space to the output labels so as to minimize the prediction error. This paper focuses on music classification based on audio signals.

2.1 Overview of Audio Features

Many audio features have been recommended in the literature for music classification. Different taxonomies exist for the classification of audio features. Scaringella [1] followed a standard taxonomy by dividing audio features used for genre classification into three groups based on timbre, rhythm, and pitch information, whereas Weihsetal [2] has categorized the audio features into another four subcategories, namely short-term features, long-term features, semantic features, and compositional features. From the perception of better understanding, Zhouyu Fu [3] divides audio features into three levels, low-level, mid-level and track-level information Fig. 2 Shows the qualities of audio features from different levels.

2.2 Overview of Low-Level Features

Low-level features play a key role in extracting useful information from a given input signal. These features can be subcategorized into two classes, namely spectral and temporal [37]. Spectral features (SFs) represent the spectral characteristics of music in a relatively short period of time. In a musical sense, SFs are used to extract the timbre or tonal characteristics of music. Timbre is always used to describe the sound; it distinguishes one sound from another sound. For example, it finds the difference in sound between a piano and guitar playing in the same note at the same volume.

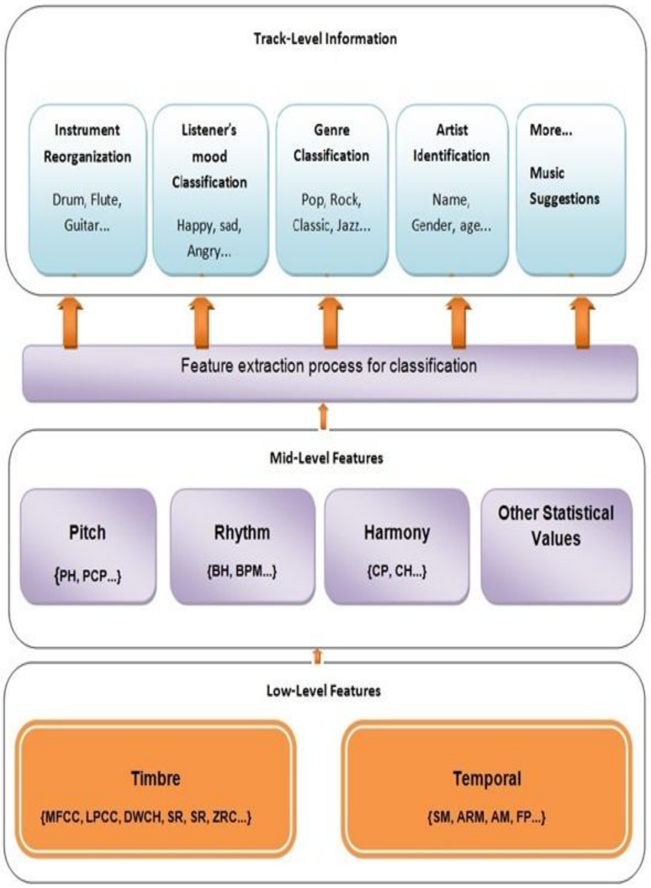

Figure 2: Qualities of audio features from Low-level, Mid-level and Track-level information.

The list of features used to characterize timbre of instruments may be found in [4]. The extraction of timbre features is closely associated with spectral analysis of the audio signal. Generally, an audio song is a bit long in duration (minutes), so the input signal for each song may contain millions of samples given high -quality frequency. Fig.3.Shows the extraction of low-level features. Instead of giving the high-quality signal directly, to improve the efficiency of a song can be split into small frames with a song duration from 10-100 ms range this process is called framing. The low-level features are extracted from the combination of both time domain and frequency domain. After framing, to extract frequency domain features spectral analysis takes place with the help of techniques like Fourier Transformation (FT), Constant Q-Transform(QT), Discrete wavelet transform (DWT), Morlet wavelet transforms (MWT) and so on.

Similarly, with the help of the zero-crossing rate (ZCR), root means square energy (RMS) [67] and crest factor (CF), time-domain features can be extracted. From the output of spectral magnitude sub band decomposition takes place to extract useful statistical information with the help of features like spectral centroid (SC), spectral roll-off (SR), spectral flux (SF) and spectral bandwidth (SB) these are also called as spectral features. Some of the spectral features used in music classification are listed in Table 1.

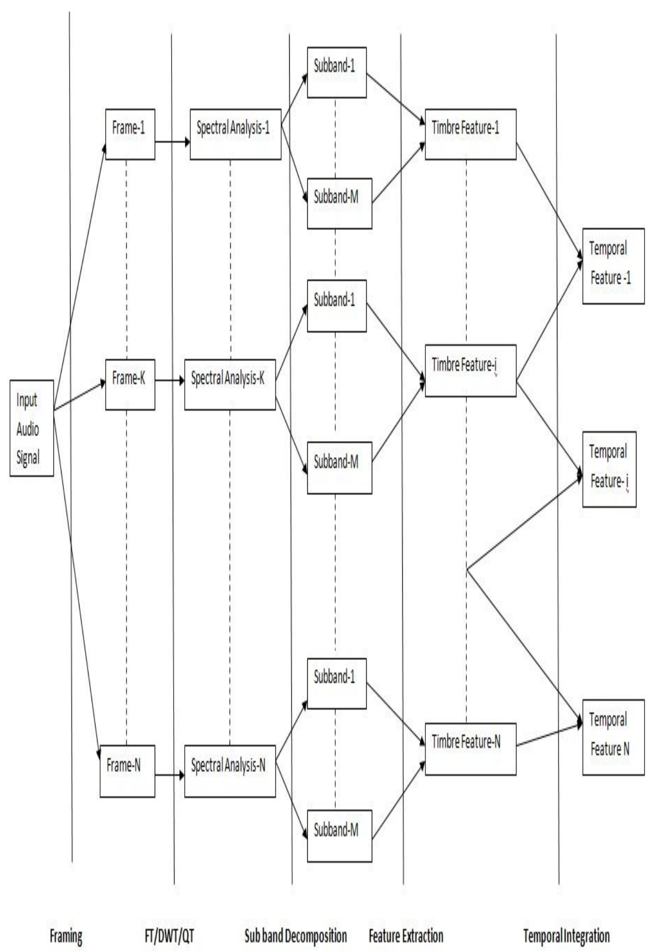

Figure 3: Step by step procedure for extracting low-level features.

Thereafter short-time Fourier transformation can be used to extract most reliable features such as Mel-frequency cepstral coefficient (MFCC), Linear predictive cepstral coefficients (LPCC), Fourier cepstrum coefficient (FCC), Octave scale cepstral coefficient (OSCC), Daubechies wavelet coefficient histogram (DWCH), Stereo panning spectrum features (SPSF), Amplitude spectrum envelope (ASE), spectral flatness Measure (SFM), spectral crest factor (SCF). For MFCC, the sub bands are linearly spaced in low frequencies are less than 1kHz and logarithmically spaced in high frequencies.

Table 1: Some of the spectral features used in Music Classification.

Class	Feature Types	Referred in
Spectral	Spectral centroid (SC)	[12],[19],[20],[23],[24]
	Spect roll-off (SR)	[12],[19],[20],[23],[24]
	Spectral flux (SF)	[12],[20],[24]
	Spectral bandwidth (SB)	[19],[20],[23],[24]
	Stereo panning spectrum features (SPSF)	[27],[28]
	Spectral flatness Measure (SFM)	[16],[17],[25]
	Spectral crest factor (SCF)	[16],[17],[25]
	Zero-crossing rate (ZCR)	[12],[19],[20],[23],[30]

The reason for using logarithms for high frequencies is to minimize the spectral resolution into lower frequencies. [36] Some of the timbre features used in music classification is listed in Table 2.

Table 2: Some of the timbre features used in Music Classification.

Class	Feature Types	Referred in
Timbre	Amplitude spectrum envelope (ASE)	[5], [6],[7]
	Constant Q-Transform(QT)	[8], [9]
	Crest factor (CF)	[10]
	Discrete wavelet transform (DWT)	[11]
	Daubechies wavelet coefficient histogram (DWCH)	[12],[13],[14],[15],[16],[17]
	Fourier cepstrum coefficient (FCC)	[18],[19]
	Linear predictive cepstral coefficients (LPCC)	[21],[22]
	Mel-frequency cepstral coefficient (MFCC)	[19],[20],[22],[26]
	Morlet wavelet transforms (MWT)	[32],[33],[34],[35]
	Octave scale cepstral coefficient (OSCC)	[6],[14],[24],[29]
	Root mean square energy (RMS)	[30],[31]

In the modern era, the music industry has made extensive use of multiple sound channels during the music recording and production stages, most of them are stereo sounds. In current music classification systems, stereo information is neglected and only the mixed sound signals are considered by adding two channels. On the other hand, temporal features (TFs) describe the relatively long-term dynamics of a music signal over time such as temporal transition or rhythmic characteristics. These include zero-crossing rate (ZCR), temporal envelope, tempo histogram, and so on as illustrated in Fig: 3, timbre features extracted from several frames, are then further integrated to extract temporal features from a given input audio signal. The extracted temporal features from a large local window are in the form of statistical parameters like mean, variance, covariance and kurtosis [20], [66]. Signal processing techniques are applied to the time-series data to extract temporal features.

The key difference is that the timbre features are extracted from a given input audio signal from a local window, whereas the temporal features are extracted from a series of extracted timbre features from a large texture window. So there is a slight increase in the computation complexity when extracting this type of temporal or time-domain features. We can apply short term Fourier transformation techniques to this time series features such as amplitude modulation (AM) [38], multivariate autoregressive models [39], probabilistic models such as hidden Markov models (HMMs) [2], [40]. HMM is a statistical model for unobservable time series data with hidden states. Similarly, diagonal-covariance Gaussians and Gaussian mixture models (GMMs) [26] are used to reduce complexity.

2.3 Overview of Mid-Level Features

Musical concepts can be best viewed and defined through human perception. It is often not enough to approximate them through a simpler concept or feature [41]. For instance, music speed is perceived as fast or as slow. It is not explained by equivalent to tempo like beats per minute. In fact, perceptual speed is better approximated by the onset rate [42]. The commonly used Mid-level features are Harmony, Pitch, and Rhythm. Such metadata could be used to improve search and retrieval, mostly helpful in query by humming and query by singing.

1. **Rhythm** is the pattern of regular or irregular tension and release in music.
2. **Pitch** is the characteristic of a sound, specifically a musical tone that is described by the frequency.
3. **Harmony** is the combination of notes of simultaneously sounded music known as chords.

The above features are easily identified by the music listeners but in contrast with system perception it is not reliable to extract such features from raw audio signal for the purpose of music analysis. Rhythm is one of the important features in audio-based music classification; it recognizes the occurrences of regular, irregular and repeated patterns in music clips. The features beat per minute (BPM) and tempo are used to describe rhythmic contents. The auto-correlation of the time-domain envelope signal is calculated. Then peaks of the auto-correlation function identify the rhythmic content of the music. Rhythmic features can be obtained such as magnitudes and locations of peaks which helps to define beat histogram (BH) [20],[43]. These rhythmic features are used for genre classification and also used for mood classification [14],[44],[45]. For instance, a slow rhythm indicates a sad song similarly, a fast rhythm indicates an energetic song.

In music mining, pitch plays an important role in applications like query-by-humming (QBH) and raga identification. Fig-4 shows the basic model of QBH.

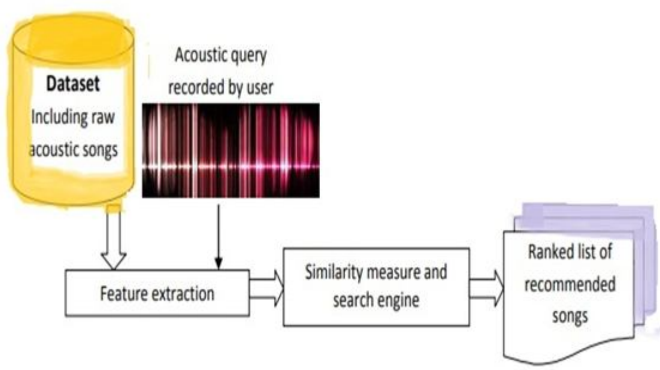

Figure 4: A typical query-by-humming based music information retrieval system.

In modern music, multiple instruments, as well as singers, are played and sung at the same time which leads to multi-pitch estimation, various algorithms are developed [46], [47] to identify the unique candidate or instrument pitches from each frame and song level pitch features using pitch histograms (PHs) [48]. The PH are used for genre classification and also used for mood classification with the help of MFCC and other perceptual features.

Another important feature is pitch class or Chroma, Each note consists of two parts, pitch, and duration. A pitch interval is the distance between two pitches, and the interval of 6, 7, 8 or 12 semitones is called an octave.

The mapping of all pitches from different octaves into a single octave in the diatonic scale. Fig-5 Shows Diatonic scale on C.

Figure 5: Diatonic scale on C, a” white note" scale.

For example, Low clef (C), Middle C, and High C all belong to the same pitch class profile (PCP) although they arehaving several variations in frequency.Fig-6 Shows Levels of Clef (C). Pitch class profile (PCP) and harmonic pitch class profile (HPCP) have been developed and widely used in melody analysis and transcription [49], [50]. The chroma feature is a simplified version of HPCP [51] and can be extracted directly from the spectrum values without attempting on pitch detection with the help of MFCC [52]. Chroma features are helpful in analyzing the melody of a song. Melody is the combination of pitch and rhythm.

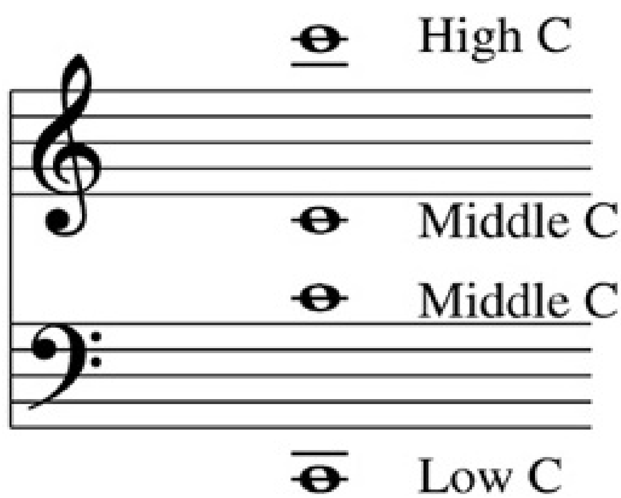

Figure 6: Levels of Clef (C).

Another important feature is harmony, which involves chords. A chord is a combination of two or more notes played simultaneously. There are several chord detection and recognition algorithms [53], [54], [55], [65]. These processes begin with pitch detection for frequencies and its partials. The PH is then compared with chord templates to find out if there's any existing chord.

Any musical sound can be categorized as vertical or horizontal components or both. Harmony majorly falls into the category of vertical aspect; it also has the horizontal aspect. Harmony in any music gives it a unique position in the audio world. however, music does exist without harmony.

To summarize the audio features, the combination of low-level and mid-level features would give better results in music processing tasks [56]. Timbre and spectral features are suitable for genre and instrument classification. Rhythm features are suitable for mood classification [14], [24],[57]. On the other hand, pitch and harmonic features are suitable for song similarity retrieval and cover song detection [49], [58], [59], [60], [61], [62], [63], [64] at a melodic level.

3. CONCLUSION

Firstly, this paper addresses the recent growth of the music industry with statistics and the importance. In this review, major research area was focused on audio features extraction procedures. Low-level and mid-level features are explained in this paper in a detailed architecture. The low-level features are timbre and temporal whereas mid-level features are Rhythm, Harmony and Pitch. These features are considered as important in order to categorize and classify the music classification tasks. Finally, the main agenda of this review is to identify the basic audio features which will be useful for better understanding of the problems and proposing optimal solutions in the music recommended applications.

REFERENCES

[1] Scaringella, N., Zoia, G., & Mlynek, D. (2006). **Automatic genre classification of music content: a survey.** *IEEE Signal Processing Magazine,* 23(2), 133-141.

[2] Weihs, C., Ligges, U., Mörchen, F., & Müllensiefen, D. (2007). **Classification in music research.** *Advances in Data Analysis and Classification,* 1(3), 255-291.

[3] Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2010). **A survey of audio-based music classification and annotation.** *IEEE transactions on multimedia,* 13(2), 303-319.

[4] Peeters, G. (2004). **A large set of audio features for sound description (similarity and classification) in the CUIDADO project**. CUIDADO IST Project Report, 54(0), 1-25.

[5] Kim, H. G., & Sikora, T. (2004, September). **Audio spectrum projection based on several basis decomposition algorithms applied to general sound recognition and audio segmentation.** *In 2004 12th European Signal Processing Conference (pp. 1047-1050). IEEE.*

[6] Lee, C. H., Shih, J. L., Yu, K. M., & Lin, H. S. (2009). **Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features.** *IEEE Transactions on Multimedia,* 11(4), 670-682.

[7] H. G. Kim, N. Moreau, and T. Sikora, **“Audio classification based on MPEG-7 spectral basis representation,”** *IEEE Trans. Circuits Syst. Video Technol.,* vol. 14, no. 5, pp. 716–725, 2004.

[8] Schörkhuber, C., & Klapuri, A. (2010, July). **Constant-Q transform toolbox for music processing.** *In 7th Sound and Music Computing Conference, Barcelona, Spain (pp. 3-64).*

[9] Brown, J. C. (1991). **Calculation of a constant Q spectral transform.** *The Journal of the Acoustical Society of America,* 89(1), 425-434.

[10] Helen, M., & Virtanen, T. (2005, September). **Separation of drums from polyphonic music using non-negative matrix factorization and support vector machine**. *In 2005 13th European Signal Processing Conference (pp. 1-4). IEEE.*

[11] Bruce, L. M., Koger, C. H., & Li, J. (2002). **Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction.** *IEEE Transactions on geoscience and remote sensing,* 40(10), 2331-2338.

[12] Li, T., Ogihara, M., & Li, Q. (2003, July). **A comparative study on content-based music genre classification.** *In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 282-289). ACM.*

[13] Li, T., & Ogihara, M. (2006). **Toward intelligent music information retrieval.** *IEEE Trans. Multimedia*, 8(3), 564-574.

[14] Yang, Y. H., Lin, Y. C., Su, Y. F., & Chen, H. H. (2008). **A regression approach to music emotion recognition.** *IEEE Transactions on audio, speech, and language processing,* 16(2), 448-457.

[15] Wang, F., Wang, X., Shao, B., Li, T., & Ogihara, M. (2009, October). **Tag Integrated Multi-Label Music Style Classification with Hypergraph**. In *ISMIR* (pp. 363-368).

[16] Benetos, E., Kotti, M., & Kotropoulos, C. (2006, May). **Musical instrument classification using non-negative matrix factorization algorithms and subset feature selection.** *In 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings (Vol. 5, pp. V-V). IEEE.*

[17] Allamanche, E., Herre, J., Hellmuth, O., Fröba, B., Kastner, T., & Cremer, M. (2001, October). **Content-based Identification of Audio Material Using MPEG-7 Low Level Description**. In *ISMIR*.

[18] Lin, C. C., Chen, S. H., Truong, T. K., & Chang, Y. (2005). **Audio classification and categorization based on wavelets and support vector machine.** *IEEE Transactions on Speech and Audio Processing,* 13(5), 644-651.

[19] Bergstra, J., Casagrande, N., Erhan, D., Eck, D., & Kégl, B. (2006). **Aggregate features and ada boost for music classification.** *Machine learning*, 65(2-3), 473-484.

[20] Tzanetakis, G., & Cook, P. (2002**). Musical genre classification of audio signals.** *IEEE Transactions on speech and audio processing,* 10(5), 293-302.

[21] Kim, Y. E., & Whitman, B. (2002, October). **Singer identification in popular music recordings using voice coding features.** *In Proceedings of the 3rd international conference on music information retrieval (Vol. 13, p. 17).*

[22] Shen, J., Shepherd, J., Cui, B., & Tan, K. L. (2009). **A novel framework for efficient automated singer identification in large music databases.** *ACM Transactions on Information Systems (TOIS),* 27(3), 18.

[23] Morchen, F., Ultsch, A., Thies, M., & Lohken, I. (2005). **Modeling timbre distance with temporal statistics from polyphonic music**. *IEEE Transactions on Audio, Speech, and Language Processing,* 14(1), 81-90.

[24] Lu, L., Liu, D., & Zhang, H. J. (2005). **Automatic mood detection and tracking of music audio signals**. *IEEE Transactions on audio, speech, and language processing,* 14(1), 5-18.

[25] Cheng, H. T., Yang, Y. H., Lin, Y. C., Liao, I. B., & Chen, H. H. (2008, June). **Automatic chord recognition for music classification and retrieval.** *In 2008 IEEE International Conference on Multimedia and Expo (pp. 1505-1508). IEEE.*

[26] Mandel, M. I., & Ellis, D. P. (2005). **Song-level features and support vector machines for music classification.**

[27] Tzanetakis, G., Jones, R., & McNally, K. (2007, September). **Stereo Panning Features for Classifying Recording Production Style.** In *ISMIR* (pp. 441-444).

[28] Tzanetakis, G., Martins, L. G., McNally, K., & Jones, R. (2010). **Stereo panning information for music information retrieval tasks.** *Journal of the Audio Engineering Society,* 58(5), 409-417.

[29] Maddage, N. C., Xu, C., & Wang, Y. (2004, August). **Singer identification based on vocal and instrumental models.** *In Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004. (Vol. 2, pp. 375-378). IEEE.*

[30] Panagiotakis, C., & Tziritas, G. (2005). **A speech/music discriminator based on RMS and zero-crossings.** *IEEE Transactions on multimedia,* 7(1), 155-166.

[31] Sell, G., & Clark, P. (2014, May). **Music tonality features for speech/music discrimination.** *In 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2489-2493). IEEE.*

[32] George, S. E. (2005). **Optical Music Recognition with Wavelets.** In Encyclopedia of Information Science and Technology, First Edition (pp. 2225-2229). IGI Global.

[33] Zhou, W., Li, J., & Shah, F. (2016, December). **Retrospect and prospect of wavelet and big data: A survey**. *In 2016 13th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP) (pp. 307-310). IEEE.*

[34] Lostanlen, V., Andén, J., & Lagrange, M. (2019). **Fourier at the heart of computer music: From harmonic sounds to texture**. *Comptes Rendus Physique, 20(5), 461-473.*

[35] Polotti P, Rocchesso D, editors. **Sound to sense, sense to sound: a state of the art in sound and music computing.** Berlin: Logos Verlag; 2008.

[36] Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2011). **A Survey of Audio-Based Music Classification and**

Annotation. *IEEE Transactions on Multimedia*, 2(13), 303-319.

[37] Jeong, I. Y., & Lee, K. (2016, August). **Learning Temporal Features Using a Deep Neural Network and its Application to Music Genre Classification.** In *ISMIR* (pp. 434-440).

[38] Pampalk, E., Rauber, A., & Merkl, D. (2002, December). **Content-based organization and visualization of music archives.** *In Proceedings of the tenth ACM international conference on Multimedia (pp. 570-579). ACM.*

[39] Meng, A., Ahrendt, P., Larsen, J., & Hansen, L. K. (2007). **Temporal feature integration for music genre classification.** *IEEE Transactions on Audio, Speech, and Language Processing,* 15(5), 1654-1664.

[40] Reed, J., & Lee, C. H. (2009, April). **On the importance of modeling temporal information in music tag annotation.** *In 2009 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 1873-1876). IEEE*.

[41] Aljanaki, A., & Aoleymani, m. (2018) **A data-driven approach to mid-level perceptual musical feature modeling**.*19th International Society for Music Information Retrieval Conference, Paris, France.*

[42] Friberg, A., & Hedblad, A. (2011). **A Comparison of Perceptual Ratings and Computed Audio Features.** *In 8th Sound and Music Computing Conference, 06-09 July 2011, Padova-Italy (pp. 122-127).*

[43] Lykartsis, A., & Lerch, A. (2015, January). **Beat histogram features for rhythm-based musical genre classification using multiple novelty functions.** *In Proceedings of the 16th ISMIR Conference (pp. 434-440).*

[44] Feng, Y., Zhuang, Y., & Pan, Y. (2003, October). **Music information retrieval by detecting mood via computational media aesthetics.** *In Proceedings IEEE/WIC International Conference on Web* Intelligence (WI 2003) (pp. 235-241). IEEE.

[45] Yang, D., & Lee, W. S. (2009, December). **Music emotion identification from lyrics.** In 2009 11th *IEEE International Symposium on Multimedia* (pp. 624-629). IEEE.

[46] Klapuri, A. P. (2003). **Multiple fundamental frequency estimation based on harmonicity and spectral smoothness.** *IEEE Transactions on Speech and Audio Processing,* 11(6), 804-816.

[47] Tolonen, T., & Karjalainen, M. (2000). **A computationally efficient multipitch analysis model.** *IEEE transactions on speech and audio processing,* 8(6), 708-716.

[48] Tzanetakis, G., Ermolinskyi, A., & Cook, P. (2003). **Pitch histograms in audio and symbolic music information retrieval.** *Journal of New Music Research,* 32(2), 143-152.

[49] Serra, J., Gómez, E., Herrera, P., & Serra, X. (2008). **Chroma binary similarity and local alignment applied to cover song identification.** *IEEE Transactions on Audio, Speech, and Language Processing,* 16(6), 1138-1151.

[50] Poliner, G. E., Ellis, D. P., Ehmann, A. F., Gómez, E., Streich, S., & Ong, B. (2007). **Melody transcription from music audio:** Approaches and evaluation. *IEEE Transactions on Audio, Speech, and Language Processing,* 15(4), 1247-1256.

[51] Ellis, D. P., & Poliner, G. E. (2007, April). **Identifying cover songs' with chroma features and dynamic programming beat tracking.** *In 2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07 (Vol. 4, pp. IV-1429). IEEE.*

[52] Ellis, D. P. (2007). **Classifying music audio with timbral and chroma features**.

[53] Gómez, E., & Herrera, P. (2004, October). **Estimating The Tonality Of Polyphonic Audio Files:** Cognitive Versus Machine Learning Modelling Strategies. *In ISMIR.*

[54] Lee, K. (2006, November). **Automatic Chord Recognition from Audio Using Enhanced Pitch Class Profile.** In *ICMC*.

[55] Chen, R., Shen, W., Srinivasamurthy, A., & Chordia, P. (2012, October). **Chord Recognition Using Duration-explicit Hidden Markov Models.** In *ISMIR* (pp. 445-450).

[56] Kitahara, T. (2010). **Mid-level representations of musical audio signals for music information retrieval. In Advances in Music Information Retrieval** (pp. 65-91). *Springer,* Berlin, Heidelberg.

[57] Yang, D., & Lee, W. S. (2004, October). **Disambiguating Music Emotion Using Software Agents.** In *ISMIR* (Vol. 4, pp. 218-223).

[58] Fang, J. T., Day, C. T., & Chang, P. C. (2017). **Deep feature learning for cover song identification.** *Multimedia Tools and Applications,* 76(22), 23225-23238.

[59] Amiripalli, S. S., & Bobba, V. (2018). **Research on network design and analysis of TGO topology.** *International Journal of Networking and Virtual Organisations,* 19(1), 72-86.

[60] Amiripalli, S. S., & Bobba, V. (2019). **Impact of trimet graph optimization topology on scalable networks.** *Journal of Intelligent & Fuzzy Systems*, 36(3),2431-2442.

[61] Tsai, W. H., Yu, H. M., & Wang, H. M. (2005, September). **Query-By-Example Technique for Retrieving Cover Versions of Popular Songs with Similar Melodies**. In *ISMIR* (Vol. 5, pp. 183-190).

[62] Bello, J. P. (2007, September). **Audio-Based Cover Song Retrieval Using Approximate Chord Sequences: Testing Shifts, Gaps, Swaps and Beats**. In *ISMIR* (Vol. 7, pp. 239-244).

[63] Amiripalli, S. S., & Bobba, V. (2019). **An Optimal TGO Topology Method for a Scalable and Survivable Network in IOT Communication**

Technology. Wireless Personal Communications, 107(2), 1019-1040.

[64] Chowdary, P. R., Challa, Y., & Jitendra, M. S. N. V. (2019, May). **Identification of MITM Attack by Utilizing Artificial Intelligence Mechanism in Cloud Environments.** In *Journal of Physics: Conference Series* (Vol. 1228, No. 1, p. 012044). IOP Publishing.

[65] Amiripalli, S. S., Kollu, V. V. R., Jaidhan, B. J., Srinivasa Chakravarthi, L., & Raju, V. A. (2020). **Performance improvement model for airlines connectivity system using network science.** *International Journal of Advanced Trends in Computer Science and Engineering*, 9(1), 789-792.

[66] Amiripalli, S. S., &Bobba, V. (2019). **Trimet graph optimization (TGO) based methodology for scalability and survivability in wireless networks.** *International Journal of Advanced Trends in Computer Science and Engineering*, 8(6), 3454-3460.

[67] Anita K, & Elizabeth, R. (2019). **Unconstraint 2-Level Cost Optimization Coding for Signal Processing.** *International Journal of Advanced Trends in Computer Science and Engineering,* 8(1), 93-100. https://doi.org/10.30534/ijatcse/2019/15812019.