Makara Journal of Technology

Volume 25 | Number 2 Article 5

8-2-2021

Music Information Retrieval Based on Active Frequency

Hardianto Wibowo
Informatic Engineering, Universitas Muhammadiyah Malang, Malang 65144, Indonesia, ardi@umm.ac.id

Wildan Suharso
Informatic Engineering, Universitas Muhammadiyah Malang, Malang 65144, Indonesia

Yufis Azhar
Informatic Engineering, Universitas Muhammadiyah Malang, Malang 65144, Indonesia

Galih Wasis Wicaksono
Informatic Engineering, Universitas Muhammadiyah Malang, Malang 65144, Indonesia

Agus Eko Minarno
Informatic Engineering, Universitas Muhammadiyah Malang, Malang 65144, Indonesia

See next page for additional authors

Recommended Citation

Wibowo, Hardianto; Suharso, Wildan; Azhar, Yufis; Wicaksono, Galih Wasis; Minarno, Agus Eko; and Harmanto, Dani (2021) "Music Information Retrieval Based on Active Frequency," *Makara Journal of Technology*: Vol. 25 : No. 2 , Article 5.
DOI: 10.7454/mst.v25i2.3977
Available at: https://scholarhub.ui.ac.id/mjt/vol25/iss2/5

Makara J. Technol. 25/2 (2021), 84–90
doi: 10.7454/mst.v25i2.3977

Music Information Retrieval Based on Active Frequency

Hardianto Wibowo[1,2*], Wildan Suharso[1], Yufis Azhar[1], Galih Wasis Wicaksono[1], Agus Eko Minarno[1], and Dani Harmanto[3]

1. Informatic Engineering, Universitas Muhammadiyah Malang, Malang 65144, Indonesia
2. Institute of Culture, Universitas Muhammadiyah Malang, Malang 65144, Indonesia
3. Aeronautical Engineering, De Montfort University, Leicester LE1 9BH, United Kingdom

[*]*E-mail: ardi@umm.ac.id*

Abstract

Music is the art of combining frequencies. A balance of frequencies gives rise to a harmonious tone. Several features of music can be analyzed, and they include sociocultural background, lyrics, mood, tempo, rhythm, harmony, melody, timbre, and instrumentation. In this study, we use the frequency of instrumentation as a feature for classification because each instrument has a frequency range. To test this frequency range, we use five music genres and one music playing skill. The five genres are *dangdut*, electronic dance music (EDM), metal, pop/rock, and reggae. The music playing skill is acoustic. Active frequencies are tested using the k-nearest neighbor method, and the results serve as basis of the accuracy of music classification. The classification accuracy for EDM, metal, and acoustic is over 70%, whereas that for dangdut, pop/rock, and reggae is less than 60%. In sum, the accuracy of music classification is influenced by the similarities in the music instruments used and the tempo.

Abstrak

Pengambilan Informasi Musik Berdasarkan pada Frekuensi Aktif. Musik adalah seni menggabungkan frekuensi. Suatu keseimbangan frekuensi menghasilkan peningkatan sampai suatu nada harmonis. Beberapa fitur musik dapat dianalisis, dan fitur-fitur tersebut mencakup latar belakang sosial budaya, lirik, suasana hati, tempo, ritme, harmoni, melodi, warna nada (timbre), dan instrumentasi. Dalam kajian ini, kami menggunakan frekuensi instrumentasi sebagai suatu fitur untuk pengklasifikasian karena masing-masing instrumen memiliki suatu kisaran frekuensi. Untuk menguji kisaran frekuensi ini, kami menggunakan lima genre musik dan satu keahlian bermain musik. Kelima genre tersebut adalah dangdut, musik tarian elektronik (electronic dance music (EDM), metal, pop/rock, dan reggae. Keahlian bermain musik tersebut adalah akustik. Frekuensi aktif diuji dengan menggunakan metode tetangga terdekat – k, dan hasil-hasilnya berfungsi sebagai basis akurasi klasifikasi musik. Akurasi klasifikasi untuk EDM, metal, dan akustik melampaui 70%, sedangkan untuk dangdut, pop/rock, dan reggae kurang dari 60%. Dalam penjumlahan, akurasi klasifikasi musik dipengaruhi oleh keserupaan dalam instrumen musik yang digunakan dan tempo.

Keywords: active frequency, music classification, music genre

1. Introduction

The creation of tones and voices and their combinations and temporal relationships result in music that features balance and unity [1]. A song is composed of tones and voices that are combined to build rhythm and harmony. The dimensions of music are divided into sociocultural background, lyrics, mood, tempo, rhythm, harmony, melody, timbre, and instrumentation.

The resemble that counts most for music's expressiveness between music's temporarily unfolding dynamic. The structure and configuration of human behavior are associated with the expression of emotions [2]. Music comes in various types, which are generally referred to as genres. Genres are defined as the grouping of music according to their similarities. Music can also be grouped according to other criteria, such as geography. A genre can be defined by musical technique, style, context, and music theme.

Music has features that can be classified into several dimensions. Thus, music can be in a one-dimensional or multidimensional form [3]-[5]. In the current work, we test one of the musical features for classification. Specifically, we test the frequency of instrumentation to determine the closeness of music genres.

Several studies have clustered music according to rhythm [3]-[6], the instruments used, or the harmonics [3]-[5]. In our work, we calculate the active frequencies of music genres. For example, electronic dance music (EDM) has a more active frequency distribution than pop music. From the patterns formed, we can classify music on the basis of active frequencies.

In calculating active frequency, music spectrum data are processed by dividing them into several bands. In this study, the raw data obtained are stored in 254 bands, which are then simplified into 5 bands consisting of low frequency for the first band and high frequency for the others.

To test the music retrieval system on the basis of active frequency, we use the k-nearest neighbor (KNN) method for similarity-based grouping. The KNN method is widely popular and has become the basis of many classification studies [7],[8]. With this method, an analysis is conducted to support future research.

2. Methods

The research methodology involves the use of active frequency, KNN method, and system scale.

Active frequency. Sound is a product of waves resulting from energy propagation. However, not all sounds can be heard by humans. Human ears can only hear sound with certain frequencies. These frequencies are divided into three groups: infrasonic frequency of 0–20 Hz, audible frequency with 20–20.000 Hz, and ultrasonic frequency > 20.000 Hz.

Audiosonic sound is the sound that humans can hear. It has a frequency range of 20–20.000 Hz. Sounds with a frequency under 20 Hz or over 20.000 Hz cannot be heard by normal human ears, but they may be audible for some people with sharp hearing. Sounds with a frequency of less than 20 Hz are known as infrasonic sounds. Those with a frequency of over 20.000 Hz are called ultrasonic sounds.

Software programs can be used to analyze audible frequency sounds. These programs are capable of recognizing sounds with a frequency range of 20–20.000 Hz. Not all sound sources have the full frequency span of 20–20.000 Hz. For instance, the sound of the human voice varies, with some sounds being heavy and some being shrill. This variation is due to human-produced sounds merely responding to certain frequencies only. Hence, some active sound frequency spans exist.

Sounds produced by musical instruments come with various frequency spans, and they only respond to certain frequency spans. As reported by The SAE Institute in www.sae.edu, sound frequency has two main parts: fundamental and harmonics.

Some tools are used to analyze digital sound files (MP3 files). Music sound files are created from the composition of one or more musical instruments. They are triggered by habituation through air or the predominant instrument frequency. For example, EDM activates all frequencies, but it is mainly dominated by longitudinal frequency range of 20–150 Hz. Humans hear such music with a high bass sound.

Table 1. Frequency-Based Instruments

Instrument	Fundamental Range	Harmonic Range
Violin	200–1.25 kHz	1.25–16 kHz
Viola	125–1 kHz	1–7 kHz
Cello	60–500 Hz	500–7 kHz
Double Bass	30–300 Hz	300–5 kHz
Piccolo	600–4 kHz	4–12.5 kHz
Flute	250–2.5 kHz	2.5–12 kHz
Oboe	225–1.5 kHz	1.5–12.5 kHz
Clarinet	150–2 kHz	2–12.5 kHz
Accordion	180–1 kHz	1–7 kHz
Alto Sax	150–800 Hz	800–8 kHz
Tenor Sax	120–650 Hz	650–8 kHz
Bassoon	60–620 Hz	620–9 kHz
Trumpet	180–1.2 kHz	1.2–5 kHz
Trombone	60–500 Hz	500–5 kHz
French Horn	60–800 Hz	800–5 kHz
Tuba	30–7 kHz	380–2 kHz
Harpsichord	40–1.5 kHz	-
Piano	25–7 kHz	-
Pipe organ	20–7 kHz	-
Keyboard	20–4 kHz	4–8 kHz
Female vocal	250–1 kHz	1–8 kHz
Male Vocal	100–800 Hz	800–8 kHz
Acoustic Guitar	80–1.5 kHz	1.5–5 kHz
Electric Guitar	81–1.5 kHz	1.5–5 kHz
Bass Guitar	40–1 kHz	1–4 kHz

Source: www.sae.edu

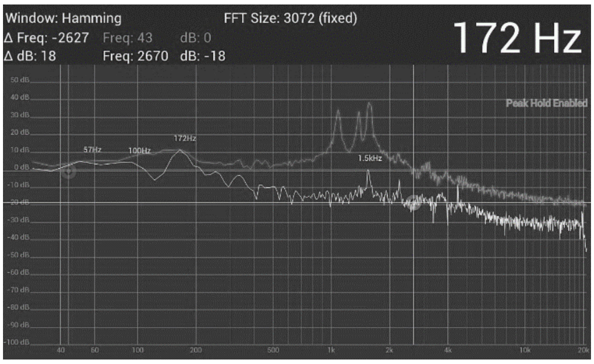

Figure 1. Software Analysis (Speeding Spectrum Analyzer) for Observing Sound File of Music

K-Nearest Neighbor (KNN). The working principle of the KNN method is to identify the shortest distance between the data to be evaluated according to the closest k neighbors in the training data. This technique includes into nonparametric classification groups. In our work, we do not pay attention to the distribution of the data to be grouped. The KNN technique is simple and easy to implement. Similar to clustering techniques, the KNN method classifies new data on the basis of their distance to multiple data/neighbors (neighbor) nearby.

The aim of the KNN algorithm is to classify new objects on the basis of the attributes and sample training. The classification does not involve any model for matching and is only based on memory. A query point is identified, and the number of objects or k (training points) closest to this query point is established. Classification is based on the most number of votes among k objects. The KNN algorithm uses the classification of neighbors as the predicted value of the new query instance. The KNN method is easy to use, and it operates on the shortest distance from the query instance to the training sample to determine the KNN.

The best k values depend on the data. A high k value reduces the effect of noise on classification, but it establishes a boundary between each classification that becomes blurry. Good k values can be selected by using parameter optimization approaches, such as cross-validation. The KNN algorithm is a unique technique as it predicts the classification on the basis of the closest training data (in other words, k = 1).

The strengths of the KNN algorithm are as follows: (1) It remains consistent when dealing with training data with much noise; (2) It is effective when the training data are huge.

Meanwhile, its weaknesses are as follows: (1) It requires the value of the k parameter (the number of nearest neighbors) to be specified; (2) Its training based on distance is not clear as the type of distance needed is not known; (3) It does not specify the types of attributes that should be used to obtain the best results; (4) Its computational cost is quite high because of the need to calculate the distance of each query instance in the whole training sample.

$$d_i = \sqrt{\sum_{i=1}^{p} \left(x_{2i} - x_{1i}\right)^2} \tag{1}$$

Scale. In the grouping calculation, we use the calculation scale for grouping frequencies that is based on the Bark scale. Scaling is a procedure to determine the location of a stimulus or response on a continuum line. Thus, scaling facilitates the production of figures on the continuum. The formula for scaling is

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} k & 0 \\ 0 & k \end{bmatrix} \begin{bmatrix} x-a \\ y-b \end{bmatrix} + \begin{bmatrix} a \\ b \end{bmatrix} \tag{2}$$

In formula dilatation with center (a, b) and the scale factor k.

Music classification. In music classification, musical features need to be grouped. These features serve as basis for establishing the similarities among music types. These features are as follows (Figure 2).

Given the dimensions of music herein, we perform our grouping on the basis of the instrumentation used in each dimension. Many approaches can be used to perform music classification. For tasks such as musical genre identification and similarity search in audio databases, audio files need to be described by suitable feature sets. As these feature sets usually try to capture diverse discriminative characteristics, acoustic representations of these feature sets should be created to support intuitive evaluation [6]. Approaches to deriving content-based audio descriptors are manifold and include the extraction of tempo, beat [9],[10], rhythm [11], pitch [12],[13], and melody [10]. In the current work, we classify music on the basis of instrumentation.

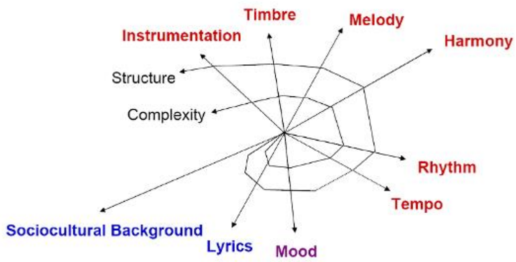

Figure 2. Dimensions of Music
Source: © ISMIR'05 Tutorial: Music Similarity, Elias Pampalk

3. Results and Discussions

Data collection. The MP3 spectrum covers the range of sounds produced by a song. Every music type has a unique spectrum that is based on the tone produced. Through this spectrum, we can observe the resulting wave model. In this research, we utilize Game Engine Unity (254 Bark scale) to observe the data spectrum produced by a song.

The data obtained are processed to determine the active frequency and then subjected to scaling into a 5 Bark scale. We utilize 30 data points for each genre, which includes dangdut, EDM, metal, pop/rock, reggae, and acoustic. The total number of data points is thus 180.

Active frequency. Active frequency is calculated to determine the activated waves. In the calculation, any noise is eliminated on the basis of noise waves with amplitudes under 10 dB. Hence, noises with amplitudes under 10 dB are eliminated.

Figure 3 shows the data processed to determine the active frequency. From such data, scaling is performed on the basis of the 5 Bark scale. The resulting data are shown in Figure 4 and 5.

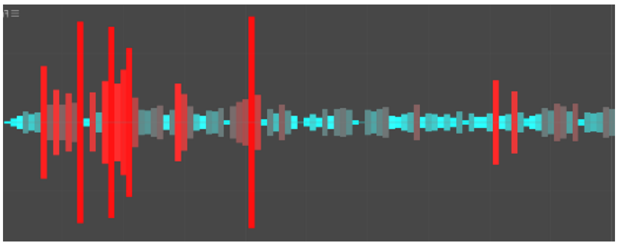

Figure 3. 254 Spectrum Music Scale

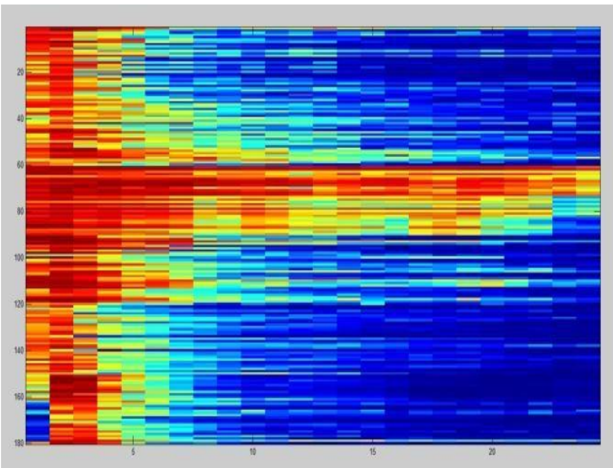

Figure 4. Data of Active Frequency with 254 Scales

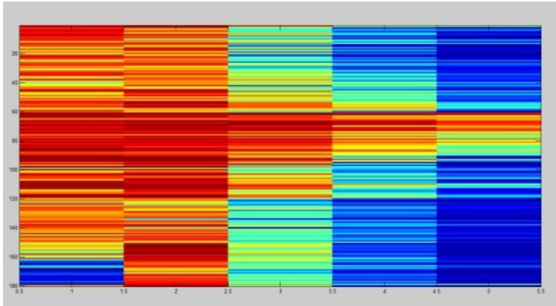

Figure 5. Result of Scaling Based on 5 Bark Scale

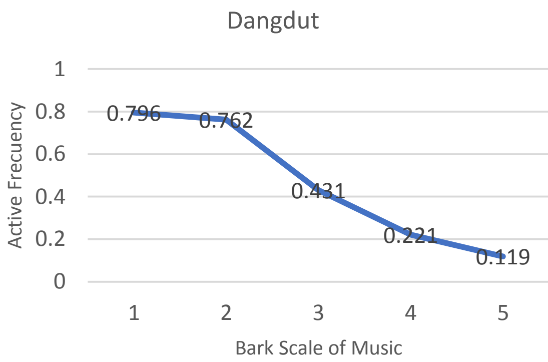

Figure 6. Average Results for Dangdut

The active frequency measurements for each genre are presented here.

Dangdut. The data on the dangdut genre have the following values:

0,796 0,762 0,431 0,221 0,119

The maximum value is 1, which means full activity. The lowest value is 0, which denotes inactivity. The data for simulation are presented in Figure 6.

Figure 6 shows the average results for dangdut. The frequencies on Bark scales 1 and 2 are more active than those on the other Bark scales. The most active bands in dangdut music are in the first and second bands, the values of which are above 0.5.

EDM. The data on the EDM genre have the following values:

0,732 0,784 0,599 0,426 0,206

The maximum and minimum values are 1 and 0, respectively. They denote the activity levels of the bands of a musical genre. The simulation results for this genre are shown in Figure 7.

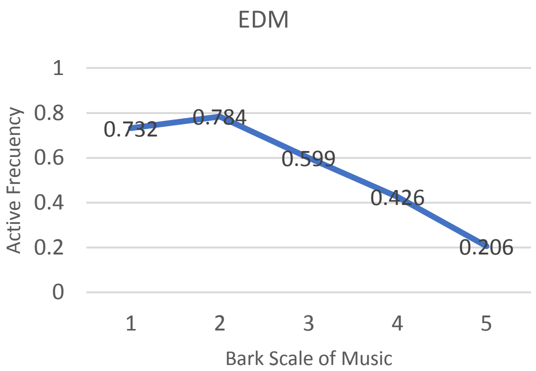

Figure 7. Average Results for EDM

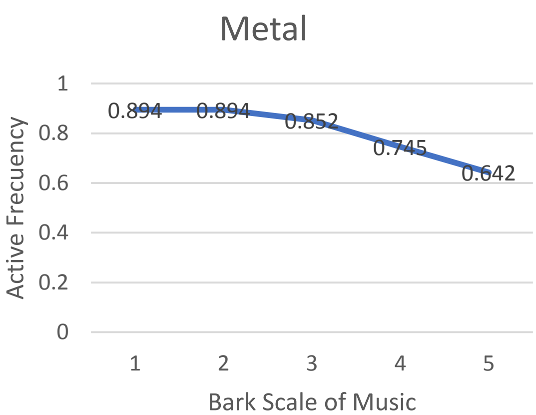

Figure 8. Average Results for Metal Genre

For the EDM genre (Figure 7), the active frequency values on bands 1 and 2 are the most dominant, with values above 0.5. For bands 3 and 4, they have a fairly large active frequency value that is close to 0.5.

Metal. Data on the metal genre have the following values:

0,894 0,894 0,852 0,745 0,642

Similarly, the maximum and minimum values are 1 and 0, respectively. The values for bands 1 to 5 are over 0.5 and thus indicate full activity. The simulation data are shown in Figure 8.

As shown in Figure 8, the active frequency is relatively even, with all frequencies being greater than 0.5.

Pop/Rock. The data on the pop/rock genre have the following values:

0,840 0,891 0,655 0,382 0,194

The dominant bands here are bands 1, 2, and 3; their values are over 0.5. The average results are shown in Figure 9.

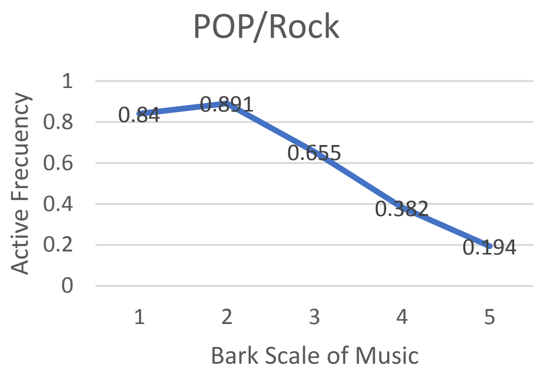

Figure 9. Average Results for Pop/Rock

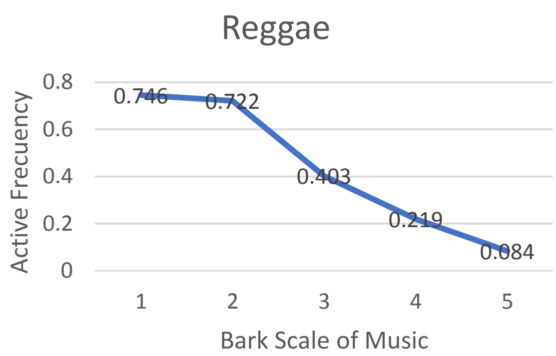

Figure 10. Average Results for Reggae Genre

The results of the pop/rock genre are almost similar to those of dangdut music. The most active frequencies are observed on Bark scales 1 and 2.

Reggae. The data on the reggae genre have the following values:

0,746 0,722 0,403 0,219 0,084

The most dominant bands are bands 1 and 2, with values over 0.5. For the fifth band, the value does not exceed 0.1, which indicates an almost absent active frequency. The average results are shown in Figure 10.

As shown in Figure 10, reggae music patterns have similarities to the patterns of pop and dangdut music. However, on the 5 Bark scale, reggae music is very lace, with an average value of less than 1.

Acoustic. The data on the acoustic genre have the following values:

0,334 0,851 0,483 0,201 0,072

Band 2 has the highest value; all the other four bands have low values below 0.5. On the Bark scale, the genre has a value below 0.1. The average results are shown in Figure 11.

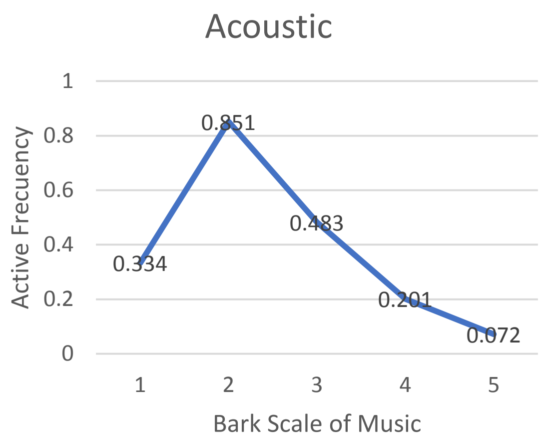

Figure 11. Average Results for Acoustic Genre

Table 2. Titles of Music used in the Test

No	Title	Author
1	Darah Biru	Megi Z
2	Blade Theme	Bryce
3	Red in Tooth and Claw	Memphis May Fire
4	Kunci Hati	Afgan
5	Traveling Home	Iba Mahr
6	Invisible	Adhitia Sofyan

As shown in Figure 11, the most active Bark scale is the second one, with the pattern of the first one being relatively low.

Testing data using KNN method. We conduct our classification by taking 60 types of songs from different genres. The test data are taken from the sample data (Table 2).

The test is performed on the basis of the data table created. The result of the test conducted from K = 1 to K = 10 is presented in Table 3.

Table 3 shows the music tested by using the KNN method. M1 to M6 refer to the music being tested, and K1 to K10 denote the classification performed using the KNN method. As shown in the Table 3, we test six genres: M1: dangdut, M2: EDM, M3: metal, M4: pop/rock, M5: reggae, and M6: acoustic. The classification results indicate that M1 and M2 are 40% and 90% similar to M4 (pop/rock), respectively. This result was obtained because in accordance with the music genre which is EDM. As for M3, it is 100% similar to metal music. Meanwhile, M4 is 40% similar to acoustic music and 30% similar to dangdut. M5 and M6 are 40% and 70% similar to the acoustic genre.

Table 3. Result of Testing by KNN

	M1	M2	M3	M4	M5	M6
K1	1	2	3	4	5	6
K2	4	2	3	1	2	6
K3	2	2	3	1	5	6
K4	5	2	3	4	2	6
K5	4	2	3	1	4	6
K6	4	2	3	6	5	6
K7	4	4	3	6	1	5
K8	5	2	3	5	4	1
K9	2	2	3	6	5	6
K10	5	2	3	6	1	4

M = Label of music used in the test (Table 2)
K = Iteration of KNN

Table 4. Precision Testing

Genre	Precision
Dangdut	10%
EDM	90%
Metal	100%
Pop/Rock	20%
Reggae	40%
Acoustic	70%

Precision. In precision testing, we calculate the level of accuracy between the information requested by the user with the answers provided by the system. In testing each genre of music consists of 30 music, from 10 times the test obtained data as follows.

As shown in Table 4, the precision for EDM, metal, and acoustic exceeds 50%. According to the observation of the active frequencies used by these types of music, we find that they have the most distinct models, hence their precision level of over 50%.

4. Conclusions

The accuracy of using active frequencies in the KNN method on the basis of human knowledge of EDM, metal, and acoustic is over 70%. The musical grouping with almost similar genres, such as dangdut, pop/rock, and reggae, is lacking because the active frequencies of the three genres are within a short distance.

From the test based on K = 1 to K = 10, we find that active frequency is a grouping of waves that are active due to musical instruments. Dangdut, pop/rock, and reggae have the closest similarity, as shown by the similarity test data. Dangdut music has 40% similarity to pop/rock than with dangdut itself when K = 10. Active frequencies are most suitable for the classification of music on the basis of the instruments used. This finding is reflected in the distribution of the test results in Table 3. Each musical instrument occupies a different frequency, and active frequency calculation is based on how often these tools bring out the vote and mutually intersect if used simultaneously. In precision testing, the highest level of accuracy is achieved for EDM, metal, and acoustic because they have the most unique patterns.

References

[1] L. Ferrara, (Ed.), Philosophy and the analysis of music: Bridges to musical sound, form and reference, Greenwood Press, 1991.
[2] S. Davies, (Ed.), Artistic expression and the hard case of pure music, 2006.
[3] C. Nopthaisong, M.M. Hasan, ICICT 2007: Proceedings of International Conference on Information and Communication Technology, 2007.
[4] C. Xu, N.C. Maddage, X. Shao, IEEE Trans. Speech Audio Process. 13 (2005) 441.
[5] Y.L. Lo, Y.C. Lin, Proceedings-2010 3rd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2010, 2010, p.112.
[6] T. Lidy, G. Pölzlbauer, A. Rauber, ICMC. (2005).
[7] G. Li, J. Zhang, 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), 2018, p.777.
[8] N.-H. Liu, Y.-H. Wu, A.L.P. Chen, Multimed. Syst. 10 (2005) 513.
[9] S. Dixon, ICMC, 2001.
[10] M. Goto, Y. Muraoka, ICMC, 1995.
[11] S. Dixon, E. Pampalk, G. Widmer, Classification of dance music by periodicity patterns, 2003.
[12] P. Lepain, J. New Music Res. 28 (1999) 296.
[13] G. Tzanetakis, A. Ermolinskyi, P. Cook, J. New Music Res. 32 (2003) 143.