Rock Genre Classification using K-Nearest Neighbor

Yoppy Sazaki[1], Adib Aramadhan[2]
Faculty of Computer Science
Department of Informatics Engineering University of Sriwijaya
Palembang, South Sumatra, Indonesia
[1]yoppysazaki@gmail.com
[2]aramadhan_adib@yahoo.com

***Abstract*— Music genre classification is a part of Music Information Retrieval. This research was a genre music detection based on signal from an audio. Divided into two processes namely extraction of features and classification. Signal would be transformed using Fast Fourier Transform to get frequency domain signal which will be processed to extract Short Time Energy, Spectral Centroid, Spectral Roll-Off, Spectral Flux, and Energy Entropy feature. Besides those features, Zero Crossing Rate would be counted from time-domain signal. in classifying phase, research using k nearest neighbor with accuracy reaching 54,44%.**

***Keywords*— Music Information Retrieval, Digital Signal Processing, Fast Fourier Transform, Feature Extraction, K-Nearest Neighbor**

I. INTRODUCTION

The development has been increasingly diverse music in which various kinds of music has sprung up, especially in rock music. Previously, only a few types of rock music are there, but now has many new emerging types of rock. Increasing the number of types of rock music certainly adds to the difficulty in determining the type of rock music. One step that can be done to overcome the problems in determining the type of rock music is to create a system that is able to classify the rock itself. There have been many studies conducted in the determination of the type of music and the famous one is MUGRAT (Music Genre Recognition by Analysis of Texture) the results of the study Karin Kosina (2002)[1]. Where MUGRAT able to classify three types of music that is classical, metal, and dance with an accuracy of classification, is the extraction and classification.

The problem of this research is how to appropriately classify the type of music, especially rock music where rock music itself is divided into several sub-genres. However, to classify a genre of music, it takes a form of input data extraction features of a musical frequencies, classification using the K-Nearest Neighbor algorithm (KNN).

The purpose of this research is as follows measuring the accuracy of the classification of the type of software usage musik rock with algoritmaK-Nearest Neighbor (KNN).

The benefit of this research is to be able to classify three types of rock music is punk rock, psychedelic rock and hard rock so it can reduce frequent debate and helping ordinary people (who are not so familiar with rock music) to be aware of the three types of rock music.

Limitation problem of this study are as follows training and audio data input wav format with 8000 Hz sample rate and only have one channel (mono), Rock type specified only 3 kinds of music that is, punk rock, psychedelic rock and hard rock. Training and input audio data is basically songs that represent the three types of rock music mentioned above. If there is someone who uses the data enter the 3 types of music outside of rock that has been determined, then the music input to be classified among the three rock music. For the training data selected band is legendary in every type of music. Sex Pistols punk rock as training data, the Jimi Hendrix Experience as training data psychedelic rock, and Black Sabbath as training data hard rock. The value of k-parameters used in the K-Nearest Neighbor classification is 5. The data used in this study is secondary data downloaded from the Internet in the form of an audio file mp3 format. Data collection techniques used in this research is to download an mp3 format songs from several internet sites which are then converted to wav format audio files with a sample rate of 8000 Hz with a single channel (mono). There are 12 songs that became the training data, where each of the types of rock music composed 4 songs. As for the input data, collected 30 songs for every type of rock music, thus totaling 90 songs.

II. K-NEAREST NEIGHBOR (KNN)

K-Nearest Neighbor (KNN) algorithm is the most commonly used in the classification, although also can be used for estimation and forecasting. KNN has several advantages, namely robustness against training data has much noise and effective when the training data is large. KNN is an example of an instance-based learning which means the training data is stored, so that the classification for unclassified new data will be compared with the training data by taking the data of the most common training (Larose, 2005)[2].

To determine the similarity, the euclidean distance function is needed to test the size of which can be used as an interpretation of proximity between two objects. Euclidean distance is represented as follows:

$$d_{euclidean}(x,y) = \sqrt{\sum_{i=1}^{N}(x_i - y_i)^2} \quad (1)$$

Where deuclidean (x, y) is the scalar distance of two vectors x and y. In the training phase, the algorithm is just doing storage feature vectors (Larose, 2005). Object in question is the object of training data and input data.

So once known each euclidean distance between the input data with training data, will be selected as the training data

parameter k has the smallest euclidean distance. Conclusion Most of the data are going to be the choice for the new classification of the input data.

Training data and the input data are selected based on the observation through an internet search engine with keywords "greatest punk rock band", "greatest psychedelic rock band", and "greatest hard rock band". The results of these observations indicate bands such as Black Sabbath, Sex Pistols, Jimmi Hendrix Experience, Pink Floyd, Jimi Hendrix Experience, and the Ramones. So for the training data represented by the Sex Pistols (punk rock), Black Sabbath (hard rock), and the Jimi Hendrix Experience (psychedelic rock)[7].

III. ANALYSIS AND DESIGN

A. Analysis of Feature Extraction

Audio files pass through the first stage of feature extraction. In the feature extraction stage, the audio file will pass through four stages: frame blocking the signal, the signal windowing, signal transformation, and counting feature. Data from the audio file a number that describes the amplitude.

1) Frame Blocking Signal Analysis

In the frame stage blocking signal, the input audio file will be divided into signal comprising several frames. The size of a frame consisting of several samples depending on how many seconds the audio file will be divided and how much sampling frequency. In this study the frame size of 50 milliseconds without overlapping. The purpose of this frame is that the blocking signal is invariant, which means no change. The outcome of this process is a signal that has been split into multiple frames.

2) Windowing Signal Analysis

Once the signal is divided into many frames, then the signal will go to the windowing stage. The purpose of windowing is to eliminate the effects of discontinuities caused by blocking frame. There are several kinds of windowing methods, the Blackman Window, Window Rectangle, and Hamming Window. In this research, windowing used is Hamming Window,

$$w(i) = 0{,}54 - 0{,}46\left(1 - cos\left(2\pi\frac{i}{n}\right)\right) \quad (2)$$

untuk i=0….n-1

Hamming window is a windowing which include having a small side lobe and main lobe of the greatest so it will be more smooth windowing results in eliminating the effects of discontinuities. The results of this stage is the signal on the frames that have been windowing..

3) Signal Transformation Analysis

At this stage it will change the original signal into a time berdomain berdomain frequency. In this research, to change the signal domain used Fast Fourier Transform (FFT). The purpose of a domain modifying the signal that the features of an input music can be calculated, where the features are the frequency characteristics of the music. All the signals that have been in the framing and windowing were going to put into the equation 3.

$$X[k] = \sum_{n=0}^{N/2-1} x[2n]W_{N/2}^{nk} + W_N^k \sum_{n=0}^{N/2-1} x[2n+1]W_{N/2}^{nk} \quad (3)$$

Having obtained the results of the real and imaginary of the signal then the next step is to enter into the equation to obtain the signal magnitude 4.

$$\mathrm{M_t[k]} = \sqrt{\mathrm{Re(X[k])^2 + Im(X[k])^2}} \quad (4)$$

.4) Calculation Analysis Features

Signal has a frequency berdomain then be computed its features. In this experiment the features that will be calculated is the Zero Crossing Rate (ZCR), Short Time Energy (STE), Spectral Centroid (SC), Spectral Roll-off (SR), Spectral Flux (SF), and the Energy Entropy (EE). But for ZCR, not used with frequency domain signal, but the signal is the original signal berdomain time.

$$ZCR = \sum_{n=1}^{N-1} |(x[n]) - (x[n-1])| \quad (5)$$

$$STE = \frac{1}{N}\sum_{n=0}^{N-1} M_t[n]^2 \quad (6)$$

$$SC = \frac{\sum_{n=0}^{N-1} M_t[n].n + 1}{\sum_{n=0}^{N-1} M_t[n]} \quad (7)$$

$$SR = 0.85 \sum_{n=0}^{N-1} M_t[n] \quad (8)$$

$$SF = \sum_{n=0}^{N-1} (N_t[n] - N_t[n-1])^2 \quad (9)$$

$$EE = -\sum_{n=0}^{N-1} p(x[n]) log_{10} p(x[n]) \quad (10)$$

B. Classification Analysis of Rock Music Type

The last phase will be the classification of types of music. The classification of music using the K-Nearest Neighbor algorithm (KNN). The features that have been previously obtained will be compared with the training data that is already available. By calculating the euclidean distance of each feature to the training data, which will be selected which has the smallest distance to the training data. From the results of euclidean distance calculation, we will get some kind of

music as a candidate k-parameters. Music input will be classified into the type of music that most of the candidates.

IV. IMPLEMENTATION AND TESTING

Rock Music classification software with Algorithm Using K-Nearest Neighbor (KNN) includes hardware, software and programming languages. The hardware used in this implementation phase is a computer with the following specifications AMD FX-4100 Quad Core 3.6 GHz, 4GB RAM, 500GB Hard Disk. The software used in the implementation of this software are Operating system Microsoft Windows 7 Ultimate, Compilers NetBeans IDE 7.2. The programming language used in the implementation of this software is Java language.

Testing of this software using black box testing method based on alignment of the conduction of software design and software implementation results. Based on the results of testing that has been done on the suitability of software design and software implementation, it can be concluded that the interface unit and built to run well.

TABLE 1
EXAMPLE TRACK 1 WITH THE NUMBER 8 FRAMES

No Frame	ZCR	STE	SC	SR	SF	EE
Frame[1]	1	32,677.24	4.4	434.21	0.0829	0.89
Frame[2]	1	30,112.23	8.7	400.98	0.0781	0.77
Frame[3]	4	35,888.7	5.1	349.87	0.0784	0.78
Frame[4]	2	39,200.64	3.3	210.21	0.0956	0.44
Frame[5]	2	40,404.66	9.4	366.63	0.0741	0.65
Frame[6]	1	29,123.41	9.1	481.46	0.0646	0.61
Frame[7]	0	56,410.87	6.8	215.52	0.0832	0.84
Frame[8]	3	49,339.21	5.0	587.29	0.0474	0.71

In order to extract the value of each individual frame into one, then

fiturZCR= std(ZCR)=1.31
fiturSTE= std(STE)/mean(STE)= 9,547.48/ 39,144.62 = 0.24
fiturSC= std(SC)= 2.35
fiturSR= std(SR)= 127.39
fiturSF= std(SF)/mean(SF)= 0.014/ 0.075= 0.19
fiturEE= std(EE)= 0.14

TABLE II
TRAINING DATA

Song	ZCR	STE	SC	SR	SF	EE	Y= Classification
1	2.42	0.5	2.34	120.88	0.18	0.13	Punk rock
2	3.05	0.44	2.77	135.44	0.17	0.16	Punk rock
3	1.54	0.57	1.98	140.55	0.17	0.14	Hard rock
4	1.76	0.21	3.47	123.32	0.2	0.15	Hard rock
5	2.25	0.36	2.14	150.6	0.21	0.11	Psychedelic rock
6	2.83	0.42	1.88	124.71	0.16	0.19	Psychedelic rock

The formula to calculate the distance euclide

$$Dk = \sqrt{\sum_{k=1}^{N}(a_k - b_k)^2} \quad (11)$$

then calculate the distance euclide one by one song, so obtained (Table III)

Obtained Punk rock category 2 pieces, 1 piece Hard rock, Psychedelic 1piece. Because most falls result in punk rock, the song expressed type 1 punk rock.

To test the classification results in the form input types of rock music 30 punk rock music, 30 music psychedelic rock, and 30 hard rock music.

TABLE III
CALCULATE THE DISTANCE

Song	D= Euclid distance	Y= classification	Including KNN category?
1	6,608	Punk rock	Yes
2	8,249	Punk rock	Yes
3	16,169	Hard rock	No
4	1,613	Hard rock	Yes
5	23,23	Psychedelic rock	No
6	3,122	Psychedelic rock	Yes

Total records were examined as many as 90 data, the sample songs that fit with the music as much as 49 types of data, sample songs that are not in accordance with the type of music as much as 41 data.

The accuracy of the classification of the type of rock music success using the K-Nearest Neighbor algorithm with the k-value of the parameter 5 = 54,44 %.

V. CONCLUSIONS AND RECOMMENDATIONS

Conclusions obtained in this research are :

1. Input audio data can be classified into a type of music rock music because of the similarity of the features of input data with training data. Type of music most of the similarity of the training data that will be input data type of music.
2. Determination of the amount of free K-parameters and the selection of training data influences the classification success percentage.
3. Based on test data as the K-90 data with the parameter 5, the results obtained success rate of 54.44% accuracy for the classification of rock music. 49 Where the data according to the type of music, and 41 the data does not match the type of music.
4. Not that too much accuracy or success due to the lack of specific criteria of features for a particular type of music, so the classification is only based similarity features the largest and closest of the input data to the training data. While the training data does not necessarily include the right to represent a type of rock music.

Advice the further development of this research include the following:

1. The number of types of music reproduction,
2. extraction features to be reproduced;
3. Reproduced audio file formats, not just * wav but can also * Mp3, * .cda, and other audio files;
4. Detection of music can be performed live (real time).

REFERENCES

[1]. Kosina.Karin, Music Genre Recognation , Fachhochschul-Studiengang, MEDIENTECHNIK UND -DE S IGN in Hagenberg, Juni 2002.

[2]. Larose. Daniel T, “Discovering Knowledge in Data: An Introduction to Data Mining”, John Wiley & Sons, Inc., Publication, 2005.

[3]. Rajinder K M, A. V. Sutagundar, S.S. Manvi, “Analysis of Automatic Music Genre Classification System” (IJCSIS) International Journal of Computer Science and Information Security. Vol. 2, No. 1,2009.

[4]. Tzanetakis. George, Cook. Perry, “Musical Genre Classification of Audio Signals”, IEEE Transactions On Speech And Audio Processing, Vol. 10, No. 5, July 2002.

[5]. Rachmat. Antonius, S.Kom,”Sound And Audio” Faculty of Informatic Engineering UniversityKristen Duta Wacana University. 2005.

[6]. Ammer.Christine, “The Facts On File dictionary of music”, Peters Edition,Ltd.2004.

[7]. Scaruffi. Piero, "A History of Rock Music 1951 - 2000”, iUniverse Inc. 2003.