Emotion Based Music Visualization with Fractal Arts

Januka Dharmapriya[1], Lahiru Dayarathne[1], Tikiri Diasena[1], Shiromi Arunathilake[1], Nihal Kodikara[1] and Primal Wijesekera[2]

[1]*University of Colombo School of Computing, Colombo 07, Sri Lanka*

[2]*University of California, Berkeley, U.S.A.*

Keywords: Emotion Recognition, Fractal Art, Information Processing, Music Emotion, Music Information Retrieval, Music Visualization, Random Forest Regression.

Abstract: Emotion based music visualization is an emerging multidisciplinary research concept. Fractal arts are generated by executing mathematical instructions through computer programs. Therefore in this research, several multidisciplinary concepts, various subject areas are considered and combined to generate artistic but computationally created visualizations. The main purpose of this research is to obtain the most suitable emotional fractal art visualization for a given song segment and evaluate the entertainment value generated through the above approach. Due to the novice nature of previous findings, limited availability of emotionally annotated musical databases and fractal art music visualizing tools, obtaining accurate emotional visualization using fractal arts is a computationally challenging task. In this study, Russell's Circumplex Emotional Model was used to obtain emotional categories. Emotions were predicted using machine learning models trained with MediaEval Database for Emotional Analysis of Music. A regression approach was used with the WEKA machine learning tool for emotion prediction. Effectiveness of the results compared with several regression models available in WEKA. According to the above comparison, the random forest regression approach provided the most reliable results compared to other models (accuracy of 81% for arousal and 61% for valence). Relevant colour for the emotion was obtained using Itten's circular colour model and it was mapped with a fractal art generated using the JWildfire Fractal Art Generating tool. Then fractal art was animated according to the song variations. After adding enhanced features to the above approach, the evaluation was conducted considering 151 participants. Final Evaluation unveiled that Emotion Based Music Visualizations with Fractal Arts can be used to visualize songs considering emotions and most of the visualizations can exceed the entertainment value generated by currently available music visualization patterns.

1 INTRODUCTION

Art, both performing and visual, is an integral part of human expression. Music, a subform of performing art, combines vocal or instrumental sounds to express human thoughts and emotions. To fully appreciate music, one must understand concepts like vocals, pitch, tone, rhythm patterns, harmonies, and timbres. Music constantly evolves over time, creating unique and intense experiences for listeners. Listeners often create mental models of the music they hear to represent their emotions, and if these models can be visually represented, it would be a remarkable achievement with many potential applications.

Existing music visualization options use various visual art forms to represent music, but they often have limitations such as focusing on only one aspect, like emotional expression or a few music features. As a result, their ability to accurately express the music's emotional expression is questionable and requires further consideration. For example, at a live musical show, the audience experiences varying levels of emotional intensity and creates imaginative visuals in their minds in response to the music. Music artists invest time and money in creating suitable background visuals using music visualization methods to enhance the show's entertainment value. However, these visuals often fail to accurately convey the music's emotional expression. Additionally, deaf individuals may not have emotive or sensorial experiences during the show. A new method that accurately represents these imaginative mental models and the true emotional expression of a music track through a visual art form would be a remarkable addition to the

Dharmapriya, J., Dayarathne, L., Diasena, T., Arunathilake, S., Kodikara, N. and Wijesekera, P.
Emotion Based Music Visualization with Fractal Arts.
DOI: 10.5220/0011929100003497
In *Proceedings of the 3rd International Conference on Image Processing and Vision Engineering (IMPROVE 2023)*, pages 111-120
ISBN: 978-989-758-642-2; ISSN: 2795-4943

currently available music visualization methods.

The best way to do this is to explore and study the commonalities between music and the visual arts. In studying those factors, we came across a lot of past research done to find the connection between musical components and mathematics. Most of these studies have confirmed (Chen et al., 2008), (Lartillot and Toiviainen, 2007), (Santos Luiz et al., 2015) that music and mathematics are more closely related to each other than they are commonly perceived to be. Also, when considering the visual arts category, we can find some images and graphics based on mathematical methods and algorithmic calculations. The best example for this is fractal arts. Fractal arts are mathematical-based algorithmic art type that is created to present calculation results of fractal objects as still images, animation, and/or media.

Studying multi-disciplinary areas of performing and visual arts reveals the interference of mathematics in both fields. Mathematics can be used as a bridge to combine divided visual and performing art forms for the same emotional expression in music visualization. In this research, we experimented to find a better mapping method between a given segment of music and one specific mathematical art form to achieve the same emotional connection between the two art forms. Fractal art was selected as the main visual art form in this research. At the abstract level, this study can be pointed out as an attempt which taken to build a relationship between multidimensional performing arts and two-dimensional visual arts (Dharmapriya et al., 2021). These kinds of musical and visual explorations expand the boundaries of computer science to generate a content creation tool to elevate the computer as a visual instrument. Music artists can then easily execute that method to generate graphic animations to express the real emotional value of their songs.

2 RELATED WORKS

When looking at the existing literature in this field, we first looked at music visualizations. Music visualization is a visual representation of specific sound features that enable users to visualize music in an understandable, meaningful, and entertaining way. Previous works show that music can be visualized by mapping music's low-level features like the size of sound, beat, and frequency spectrum to correlate with the shapes, colours, etc. of images (Ciuha et al., 2010), (Lee and Fathia, 2016). Lee and Fathia present an interactive theme-based design of a music player using Processing. Colours and player control functions are used to interact with the contents by changing the brightness of the colours to distinguish the music's different moods. Ciuha, Klemenc, and Solina propose a method to map colours to concurrent music tones. Both of these music visualization methods use a limited amount of music features and graphical features. Another possible approach for music visualization is combining a music track with a series of photos based on the emotions expressed in the music track (Chen et al., 2008). By considering emotions, Chen et al. have trained two machine learning models separately for music and image features and then mapped them together to show a slideshow of photos for each music track.

Similar to music visualization, emotion visualization is also a research area that uses mapping between emotions and graphical features such as colours as a basis (Whiteford et al., 2018), (Bialoskorski et al., 2009a). Bialoskorski, Westerink, and Broek introduce an affective interactive art system, Mood Swings, which interprets and visualizes emotions expressed by a person using colours.

A potential benefit of music visualization is that the deaf, deafened, and hard of hearing community can also enjoy music through the visualization (Fourney and Fels, 2009). Fourney and Fels present a study that has used visualizations created from three different visualization tools, and then evaluated by deaf, deafened, and hard-of-hearing participants in a focus group setting.

Some previous studies focus on the reverse process of music visualization, which is image musicalization. Margounakis and Politis present some fundamental concepts of composing a musical piece's melody by analyzing an image's colour values (chromatic synthesis) (Margounakis and Politis, 2006). Zhao et al. use Principles-of-art-based emotion features (PAEF), which are the unified combination of representation features derived from different principles, including balance, emphasis, harmony, variety, gradation, and movement as a basis for image musicalization (Zhao et al., 2014a), (Zhao et al., 2014b).

As most of these music visualizations use emotion-based mappings, we also looked at the relationship between music and emotions. Music emotion recognition is a well-researched area that includes many frameworks for detecting emotions in music, the most noted being Russell's two-dimensional Valence-Arousal (V-A) model (Russell, 1980). There are many other emotion models such as categorical and dimensional psychometric models which can be effectively used for analyzing the emotions expressed by a piece of music (Kim et al., 2010), (Li and Ogihara, 2003), (Sorussa et al., 2020a), (Fagerberg et al., 2004), (Yang and Chen, 2012), (Yang et al., 2008b). Most of these researches have used a few music features such as timbre, tonality, rhythm, etc.. They have classified songs or music clips according to the emotion they evoked using different emotion models.

We also reviewed the literature regarding fractal arts and found that it is a limited research area. Even though there are studies that show how useful fractal arts can be for reducing stress and anxiety (Taylor, 2006), (Averchenko et al., 2017), there could not find researches which have used fractal arts for music visualization purposes.

3 MUSIC EMOTION DETECTION

3.1 Emotion Model Selection

In music psychology, we cannot find specific details and a standard music mood taxonomy system about the basic emotions that music can convey to listeners. However, there are several fundamental psychological studies that have focused on different emotions that humans can perceive (Russell, 1980), (Thayer, 1989). Those studies' proposed emotional models can be categorized into two main approaches: *categorical and dimensional psychometric*.

Categorical psychometrics consider the disrupted perception of emotions. This approach can be used to cluster emotions into more classes or clusters. Hevner's adjective checklist (Hevner, 1935) is one of the main emotional models categorized under this approach. This emotional model consists of 67 adjectives organized under eight clusters circularly. Eight clusters are formulated by grouping similar adjectives into a related emotional group by depicting the high inter-cluster similarity. When compared to the dimensional psychometric, the categorical schematics can be easily understood. However, the problem is that some emotional adjectives differ from one language to another while having different meanings. This ambiguity will be problematic if we adopt this approach as a main emotional categorization approach since there are confirmed facts on past research that had problems in finding definitive ways to discriminate adjectives that have closer meaning (Yang and Chen, 2012), (Yang et al., 2008a).

Dimensional psychometrics considers multi-dimensional space to plot the fundamental emotions rather than clustering them into several uni-polar conditions or clusters. The circumplex model of affect proposed by Russell (Russell, 1980) consists of the two bipolar dimensions that affect mood responses: arousal and valence. These two dimensions are represented as a valence-arousal (VA) plan where the X-axis shows the negative valence to positive valence (also known as negative to positive pleasant) and the Y-axis shows the negative to positive arousal. Furthermore, Russell's model comprising of a list of 28 adjectives that are located in eight categories of VA plan, as shown in Fig. 1.

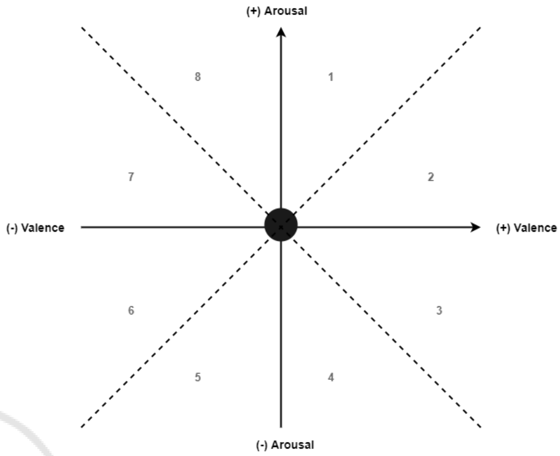

Figure 1: Illustration of Russell's circumplex model of affect with the eight emotional categories (Russell, 1980).

The main advantage of adopting Russell's dimensional psychometric is reducing several mood patterns in the categorical psychometric into the two dimensions. It will be of great importance when coming to the computational categorization of music emotions in our research. Because of that, Russell's circumplex model of affects is adopted as the basis music emotional model of our research. Furthermore, the VA plane was divided into eight emotional categories while following the logic proposed by Sorussa, Choksuriwong, and Karnjanadecha (Sorussa et al., 2020b), as shown in Table. 1.

3.2 Music Database Selection

There are many music-related databases, but very few musical databases relevant to Computer Science and Information Processing studies. Especially if the machine learning model is to be used considering emotional annotations. We recognized three main data sources (1000 Songs Emotional Database (Soleymani et al., 2013), MER500 (MakarandVelankar, 2020), The MediaEval Database for Emotional Analysis of Music (DEAM) [(Alajanki et al., 2016) & (Eyben et al., 2013)]) relevant to our study. However, due to the results showed in previous studies [(Sorussa et al., 2020c) & (Aljanaki et al., 2017)] we selected the DEAM data set.

Table 1: Basic logic for the eight emotional classes of VA Plan.

Category	VA logic range	Emotions
Category 1	High Arousal & Positive Valence $(Valence) \leq (Arousal)$	Aroused, Astonished, Excited
Category 2	High Arousal & Positive Valence $(Valence) > (Arousal)$	Delighted, Happy
Category 3	Low Arousal & Positive Valence $(Valence) \geq (\lvert Arousal \rvert)$	Pleased, Glad, Serene, Content, At Ease, Satisfied, Relaxed
Category 4	Low Arousal & Positive Valence $(Valence) < (\lvert Arousal \rvert)$	Calm, Sleepy
Category 5	Low Arousal & Negative Valence $(\lvert Valence \rvert) \leq (\lvert Arousal \rvert)$	Tired, Droopy, Bored
Category 6	Low Arousal & Negative Valence $(\lvert Valence \rvert) > (\lvert Arousal \rvert)$	Depressed, Gloomy, Sad, Miserable
Category 7	High Arousal & Negative Valence $(\lvert Valence \rvert) \geq (Arousal)$	Frustrated, Distressed
Category 8	High Arousal & Negative Valence $(\lvert Valence \rvert) < (Arousal)$	Annoyed, Afraid, Angry, Tense, Alarmed

In this study, only 45s song segments were considered. Then dynamic annotations (Arousal, Valence) related to those songs were summarized to mean value. Arousal/ valence value is then combined as the target label with the extracted music features relevant to the song.

3.3 Feature Extraction

Similar to music databases, a limited number of tools are available to extract music features computationally. MARSYAS, MatLab MIR toolbox, and JAudio (McEnnis et al., 2005) are currently available software tools. According to a previous study, (Abeysinghe, 2016) all default features marked in JAudio (as shown in table 2) are used to extract music features. Each feature is extracted for the average value and standard deviation value. There were 72 features in total (14 features, 32 dimensions on average, and standard deviation).

After extracting music features, they were pre-

Table 2: Extracted Music Features from the DEAM Data set.

Feature Name	Dimension
Spectral Centroid	1
Spectral Rolloff Point	1
Spectral Flux	1
Compactness	1
Spectral Variability	1
Root Mean Square	1
Fraction of Low Energy Windows	1
Zero Crossings	1
Strongest Beat	1
Beat Sum	1
Strength of Strongest Beat	1
LPC	10
Method of Moments	5
Area Method of Moments of MFCCs	10

possessed using Waikato Environment for Knowledge Analysis software (WEKA). Firstly, extracted features were normalized and then standardized. An automatic feature selection filter was used to extract the most suitable features. For Arousal, there were 13 features, and there were 09 features for valence.

3.4 Classification Approaches

As mentioned previously, after prepossessing the extract music features, we conducted 07 subsequent experiments separately for both emotional dimensions using the WEKA tool. To show more detailed classification results of those experiments, Table. 3 and 4 show the results of Correlation coefficient (R), Mean absolute error (MAE), Root mean squared error (RSE), Relative absolute error (RAE) and, Root relative squared error (RRSE) among the different classifiers, where Table. 3 corresponds to the regression of arousal Table. 4 regression of valence.

From Table. 3, the highest correlation coefficient of 0.817 was delivered when using the “Random Forest” classifier, and the lowest correlation coefficient of 0.580 was obtained when using the “Decision Stump” classifier. Overall, these results indicate the “Random Forest” classifier has good performance in predicting the arousal values. Also, it has comparatively less MAE, RSE, RAE, and RRSE values compared with the other classifiers.

From Table. 4, the highest correlation coefficient of 0.602 was obtained when using the “Random For-

Table 3: Experimental results of the prediction of Arousal.

	R	MAE	RSE	RAE	RRSE
Random Forest	0.817	0.125	0.163	52.224%	57.714%
REP Tree	0.730	0.149	0.195	62.429%	68.721%
Linear Regression	0.810	0.126	0.166	52.971%	58.586%
Multi layer Perceptron	0.811	0.158	0.199	66.220%	70.182%
Simple Liner Regression	0.6138	0.1757	0.224	73.378%	78.942%
Decision Stump	0.580	0.183	0.231	76.801%	81.473%
Random Tree	0.678	0.166	0.222	69.628%	78.351%

Table 4: Experimental results of the prediction of Valence.

	R	MAE	RSE	RAE	RRSE
Random Forest	0.602	0.15	0.190	75.102%	79.853%
REP Tree	0.520	0.160	0.205	80.528%	85.841%
Linear Regression	0.575	0.152	0.195	76.496%	81.804%
Multi layer Perceptron	0.558	0.180	0.229	90.184%	95.832%
Simple Liner Regression	0.398	0.173	0.219	87.095%	91.844%
Decision Stump	0.377	0.178	0.221	89.289%	92.625%
Random Tree	0.381	0.210	0.266	105.225%	111.659%

est" classifier, and the lowest correlation coefficient of 0.377 was obtained when using the "Decision Stump" classifier. The accuracy of each classifier used for valence prediction is low compared to the same classifiers used to predict arousal values. However, previous studies have shown that prediction results of arousal have been always more accurate than the valence prediction (Yang et al., 2008), (Fukayama and Goto, 2016), (Weninger et al., 2014), (Nguyen et al., 2017). Therefore, we assumed that this relatively low correlation coefficient was acceptable for our study.

Overall, these results show that the "Random forest" classifier performs well in predicting arousal and valence values, and its error values are relatively low compared to other classifiers. Thus, we used the "Random Forest" classifier as the main classifier for our study to obtain arousal and valence values.

4 VISUALIZING MUSIC EMOTIONS USING FRACTAL ART BASED ANIMATIONS

The JWildfire software was used in this study as the main fractal art generation tool. Currently, JWildfire software supports generating 50 types of different image types including the most commonly used two fractal art types of "Mandelbrot" and "Julia set (also known as Julians)". The "Mandelbrot" and "Julia Set" fractals are produced using the same formulas. But those two forms have different starting values. Because of that, if we can map one of these fractal arts to the emotional values of a music segment, it can easily adapt that to other types as well. Therefore, we have selected "Julia set" as the main focused fractal art type in the first phase of our research study with the aim of continuing this for the "Mandelbrot" fractal as well.

When studying the "Julians" fractal generation file which was in the source code of the JWildfire application, we identified that it consists of main three transformations. These transformations consist of separate values for position, colour, weight, and variations. Furthermore, we found that it is possible to map colour and positioning values of the fractal art by considering predicted valence and arousal values which we received from above mentioned computational models as described in the below sections.

4.1 Processing of Color Values in Fractal Art

Colour was chosen as a primary mapping function between the music segments and visual arts by considering the arousal and valence values obtained from the above-mentioned computational process. The reason for choosing colour as the primary mapping function is a strong relationship between colours and emotions. In any art medium, artists are using colours to excite the emotional values in their audience by considering the meaning of colours with associated factors such as personal experiences, cultural factors, and evolution (L., 2005). Because of that choosing a colour to demonstrate a specific emotion can be a very subjective and doubtful topic.

However, there are several fundamental past research works on the relationship between colours and emotions (Ståhl et al., 2005), (Bialoskorski et al., 2009b). The EMoto: an emotional text messaging service developed by Ståhl, Sundström, and Höök (Ståhl et al., 2005) provides a chance for users to adjust the affective background that includes colours, shapes, and animations of a text message to fit the emotional expression that the user wants to achieve. In this study, colours and emotions are mapped by following Itten's circular colour model (Itten, 1974) where red colour is used to represent high arousal, and blue colour is used to represent low arousal. Also, the mood swing system proposed by Bialoskorski, Westerink, and Broek, which interprets and visualizes the affect expressed by a person, has used Itten's transformed colour circle with six colour combinations to reflect the emotional state of the user, based on the user's movements (Bialoskorski et al., 2009b). In this study, red and orange colors represent high arousal, while blue and green colors represent low arousal.

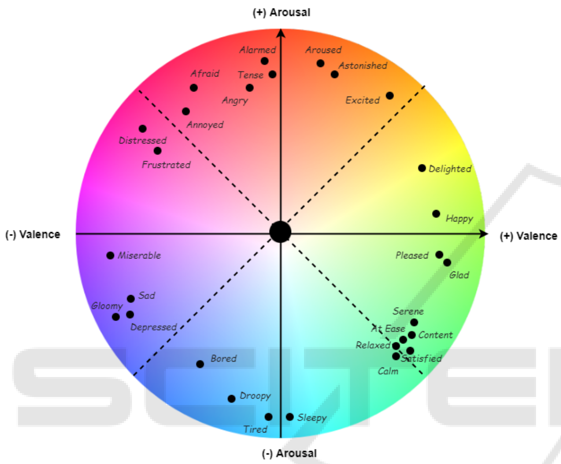

Figure 2: Illustration of mapping of the Itten's colour system (Itten, 1974) to Russell's circumplex model of affect(Russell, 1980).

In our approach, we also adopted Itten's circular colour model as the basis colour model, and then adjusted it to fit Russell's circumplex model of affect, as is shown in Fig. 2. When the arousal and valence are neutral, the colours fade to white towards the circle's center. To represent high arousal, a combination of red-orange colours is used, and to represent low arousal; blue colour is used by following the theories and studies mentioned above.

As we get two values corresponding to Valence and Arousal, we first tried fitting them as (x, y) coordinates in the colour wheel given in Fig. 2 and getting the RGB value of the specific point using Java code. A downside of directly mapping the coordinates was that colours we get as output were very light due to most of the songs having low Valence and Arousal values. As it is clear from Fig. 2, the colours near the center of the wheel tend to be lighter. Since we need distinct colours to clearly depict the difference in emotions evoked by different songs, we re-scaled the values of valence and arousal when using them as input to the Java code. By slightly increasing the input values from this method we managed to get colour readings as RGB values that are distinct to each emotion. This initially received RGB colour values were assigned to the 2nd transformation in the fractal art.

Color value (light colour related to the predicted emotional colour) for the transformation 1 assigned considering $\frac{r}{3}$ circle's perimeter (point A in Fig. 3) in the colour wheel. Colour value for transformation 3 (bright colour related to the predicted emotional colour) was obtained from the perimeter of the color wheel (point B in Fig. 3). Fig.3 shows the diagram related to this calculation. Improvements to the predicted colour were added as the same colour variations to the original predicted colour. Angle with x-axis calculated using equation 1.

$$\theta = \tan^{-1}(\frac{x}{y}) \tag{1}$$

Color coordinates (x", y") relevant to the transformation 1 obtained using equations 2 and 3. (division by 255 was performed to map the coordinates with the colour wheel)

$$x'' = \frac{\cos(\theta) \times 85}{255} \tag{2}$$

$$y'' = \frac{\sin(\theta) \times 85}{255} \tag{3}$$

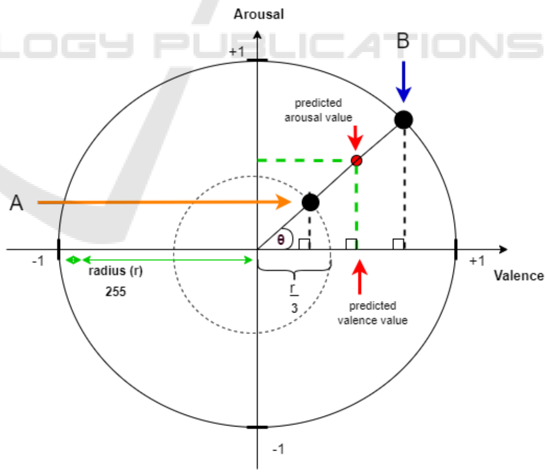

Figure 3: Colour wheel diagram for colour coordinates calculation.

Color coordinates (x"', y"') relevant to transformation 3 were obtained using equations 4 and 5. (division by 255 was performed to map the coordinates with the colour wheel)

$$x''' = \frac{\cos(\theta) \times 255}{255} = \cos(\theta) \times 1 \tag{4}$$

$$y''' = \frac{\sin(\theta) \times 255}{255} = \sin(\theta) \times 1 \qquad (5)$$

4.2 Arranging Positioning Values in Fractal Art

We set the position values of each transformation to give a unique starting point for each Julian fractal based on the music clip's predicted valence and arousal values. However, there was an issue due to incompatibility between JWildfire's graphic coordinate system and Russell's circumplex model. The graphic coordinate system used in the JWildfire application is slightly different from the standard geometric cartesian coordinate system followed in Russell's model. The main difference was the orientation of the y-axis is upside down on the graphic coordinate system used in the JWildfire application. Also, the maximum and minimum values of the circumplex model differ from the JWildfire application.

Because of that, we followed 2D Transformation rules for scaling and rotating for coinciding with both coordinate systems. Scaling factors for each axis of the JWildfire application were calculated by dividing the length of the relevant axis in the JWildfire application by the length of the relevant axis in the circumplex model. Assuming the predicted valence and arousal values as initial coordinates (x, y) and scaling factors (Sx, Sy), we can derive new coordinates (X', Y') to adjust the positioning point of the JWildfire application using the following equations.

$$X' = X \times Sx = 3X; (Sx = 3) \qquad (6)$$

$$Y' = -1(Y \times Sy = -2Y; (Sy = 2) \qquad (7)$$

Example 1. If the valence and arousal are predicted as -0.27 and -0.068 through the built emotional prediction models, the initial RGB value will be calculated as 202, 166, 225. This initial RGB value is set to the input colour value of the 2^{nd}transformation in the fractal art. According to the above mentioned colour mapping, two shade values of the initial RGB value will be calculated as 196, 157, 228 and 133, 85, 234. Those RGB colour values are set to the 1^{st} and 3^{rd} transformations in the fractal art. New coordinates values will be calculated as the x-axis equal to -0.81 (By using equation number 1) and the y-axis equal to 0.136 (By using equation number 2). From these values, we can generate different images (accordingly 25 images at once) which have the same colour patterns and unique coordinates positioning as shown in Fig.4.

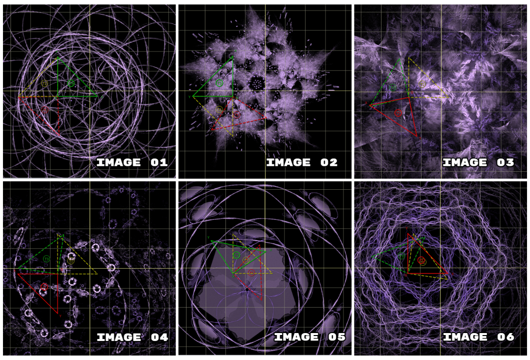

Figure 4: Sample images for same valence and arousal values.

4.3 Fractal Arts with Background Image

As Fig. 5 shows this is an enhanced feature to highlight the predicted emotion. In order to improve the emotional value given by the visualization, a background image relevant to the emotion used a gray scale. Gray-scale was used to give more priority and highlight the emotional colour. The transparency level of fractal art is also configured to the maximum level. To achieve this, an external image-providing API was combined with the JWildfire.

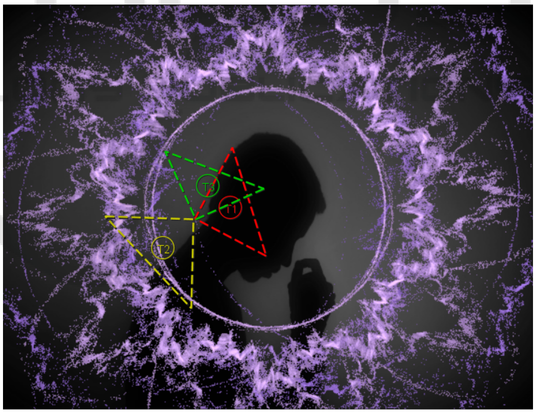

Figure 5: Fractal Art after assigning background image.

4.4 Animating Fractal Art According to the Song Variations

After creating fractal art according to the process described in the above section, the final step is to arrange fractal art animations. In JWildFire there are Global Scripts and XForm Scripts to animate a given fractal art. Global scripts animate fractal art as a whole. XForm Scripts animate fractal art considering transformations. In this study, Rotate roll was selected as the only global script. Other scripts are more suitable for 3D fractals and some scripts move away the fractal from the screen for a long time period. Due to that reason, other scripts are not considered in this study.

Table 5: Configurations used to render fractal image series.

Configuration Name	Value
Frame rate	12 frames per second
Duration	540 frames
Total video duration	$\frac{540}{12}$ = 45 seconds
Resolution	800 x 600
Quality	very low quality

For the XForm Scripts, Rotate_2ND_XForm (with music variations), Rotate_Final_XForm, Rotate_Post_Full (with music variations), Rotate_Post_2ND_XForm (with music variations) were selected. Music variations are assigned to some selected scripts. As Fig. 6 shows, it synchronizes the fractal art according to the motion curve related to the song.

In order to create fractal visualization, fractal images were generated as a series of PNG images. Table 5 shows the configurations used to render the fractal image series. After creating image series those were combined with the relevant song segment considering the same FPS rate in VirtualDub (free and open-source software) video processing utility tool.

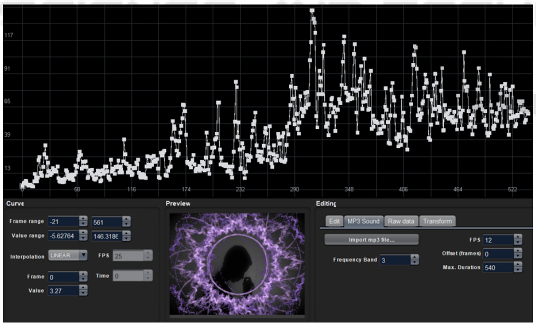

Figure 6: Motion curve for a XForm script.

5 RESULTS & FINDINGS

The emotional perception of acoustic and visual media is a very subjective topic. Hence, it is difficult to evaluate the final research outcomes through an objective evaluation. Hence, we evaluated our research outcomes through a subjective user evaluation. The final evaluation was conducted as an online survey. For that, 12 song segments (45-second duration) were selected from languages of English, Sinhala, Tamil, and Hindi. When selecting songs, at least one song segment was selected to represent each category of Rusell's circumplex emotional model as shown in Table 6. For each selected song segment, 03 types of music visualization options were created. In addition to the one generated by our research, we generated another random visualization by using the JWildfire application without considering emotional perceptions and mathematical relationships. As a third one, we used an animation generated using Windows Media Player. Then the created 03 types of visualizations for one song segment were combined as one sequence in a video. All of the combined videos were uploaded to a YouTube channel and divided into 03 playlists by attaching a separate questionnaire form for each playlist.

Table 6: Details of the song segments used in the final evaluation.

No	Song	Duration (s)	Lang.	Cat.
01	Uptown Funk	0.05 - 0.50	English	1
	Naraseeha Gatha	1.49 - 2.34	Sinhala	6
	Sunn Raha Hai	1.00 - 1.45	Hindi	7
	Muqabla	0.06 - 0.51	Tamil	1
02	Aradhana	1.15 - 2.00	Sinhala	4
	Singappenney	0.42 - 1.27	Tamil	2
	My Immortal	1.45 - 2.30	English	6
	Final Countdown	3.50 - 4.35	English	8
03	Photograph	0.10 - 0.55	English	5
	Kasthuri Suwandaki	0.49 - 1.34	Sinhala	1
	Yenti Yenti	0.57 - 1.42	Tamil	3
	Chand Chhupa	0.30 - 1.15	Hindi	2

We managed to get a total of 604 responses from 151 respondents aged from 18 to 65. Since there were 3 Google forms with each of them having 4 songs, we managed to get around 50 responses for each song. Microsoft Excel was used to pre-process, analyze, and visualize the data gathered from these surveys. We focused on comparing our visualization with the other 2 based on 4 different criteria. The following questions were asked to cover those criteria.

- Question 1 (emotion): Which visualization gives a feeling similar to the emotions of the song?
- Question 2 (colour): Which visualization's colors are most relevant to the emotional feeling of the song?

- Question 3 (synchronization): In which visualization is the music most synchronized with the animation?
- Question 4 (entertainment value): Which visualization is the most suitable as a background visual in a musical show?

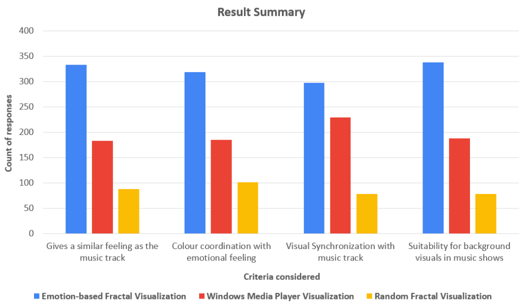

Figure 7: Comparison of different visualizations.

Fig. 7 shows the summary of the evaluation results under the 4 main evaluation criteria considered. It is clear from this graph and results analysis that our emotion-based fractal visualization is the most preferred visualization while Windows Media Player visualization and random fractal visualization rank second and third respectively for every category. Also, by analyzing the correlation between demographic data gathered and answers to these questions, it was found that these visualization preferences do not change much with demographic changes such as gender, age, language, etc. and they are much more dependent on personal preference.

6 CONCLUSION

This paper uses fractal art-based animations to visualize music. The emotional model used is Russell's circumplex model of affects, which measures Arousal and Valence. The approach uses the MediaEval DEAM database for acoustic data, with 72 acoustic feature values extracted per 45s song segment using JAudio. Different experiments were conducted using WEKA to find the best classifier for predicting arousal and valence values. The "Random Forest" classifier was found to be the most accurate. This study focuses on using "Julia set" fractals generated by JWildfire to visualize music by mapping colours and positioning values based on predicted arousal and valence values. A relevant background image is used for emotional expression and synchronization with the music. However, there is room for improvement, and evaluation was limited due to Covid-19 restrictions. The system could potentially be used to reduce stress and anxiety. This approach is a preliminary attempt to create a suitable music visualization method and could inspire more research on music visualization.

REFERENCES

Abeysinghe, H. (2016). *Automatic Genre Classification for Sri Lankan Music with Machine Learning Approaches*. 1 edition.

Alajanki, A., Yang, Y.-H., and Soleymani, M. (2016). Benchmarking music emotion recognition systems. *PLOS ONE*. under review.

Aljanaki, A., Yang, y.-h., and Soleymani, M. (2017). Developing a benchmark for emotional analysis of music. *PLOS ONE*, 12:e0173392.

Averchenko, A., Korolenko, P., and Mishin, A. Y. (2017). Optical-physical aspects of fractal art therapy. In *2017 Progress In Electromagnetics Research Symposium-Spring (PIERS)*, pages 3418–3421. IEEE.

Bialoskorski, L. S., Westerink, J. H., and van den Broek, E. L. (2009a). Mood swings: An affective interactive art system. In *International conference on intelligent technologies for interactive entertainment*, pages 181–186. Springer.

Bialoskorski, L. S. S., Westerink, J. H. D. M., and van den Broek, E. L. (2009b). Mood swings: design and evaluation of affective interactive art. *New Review of Hypermedia and Multimedia*, 15(2):173–191.

Chen, C.-H., Weng, M.-F., Jeng, S.-K., and Chuang, Y.-Y. (2008). Emotion-based music visualization using photos. In *Proceedings of the 14th International Conference on Advances in Multimedia Modeling*, MMM'08, page 358–368, Berlin, Heidelberg. Springer-Verlag.

Ciuha, P., Klemenc, B., and Solina, F. (2010). Visualization of concurrent tones in music with colours. In *Proceedings of the 18th ACM International Conference on Multimedia*, MM '10, page 1677–1680, New York, NY, USA. Association for Computing Machinery.

Dharmapriya, J., Dayarathne, L., Diasena, T., Arunathilake, S., Kodikara, N., and Wijesekera, P. (2021). Music emotion visualization through colour. In *2021 International Conference on Electronics, Information, and Communication (ICEIC)*, pages 1–6.

Eyben, F., Weninger, F., Gross, F., and Schuller, B. (2013). Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM '13, pages 835–838, New York, NY, USA. ACM.

Fagerberg, P., Ståhl, A., and Höök, K. (2004). emoto: emotionally engaging interaction. *Personal and Ubiquitous Computing*, 8(5):377–381.

Fourney, D. W. and Fels, D. I. (2009). Creating access to music through visualization. In *2009 IEEE Toronto International Conference Science and Technology for Humanity (TIC-STH)*, pages 939–944. IEEE.

Fukayama, S. and Goto, M. (2016). Music emotion recognition with adaptive aggregation of gaussian process regressors. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 71–75.

Hevner, K. (1935). Expression in music: A discussion of experimental studies and theories. *Psychological Review*, 42(2):186–204.

Itten, J. (1974). The art of color: The subjective experience and objective rationale of color.

Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, B. G., Richardson, P., Scott, J., Speck, J. A., and Turnbull, D. (2010). Music emotion recognition: A state of the art review. In *Proc. ismir*, volume 86, pages 937–952.

L., Z. V. (2005). The expressions of colours. In *DiGRA &#3905 - Proceedings of the 2005 DiGRA International Conference: Changing Views: Worlds in Play*.

Lartillot, O. and Toiviainen, P. (2007). Mir in matlab (ii): A toolbox for musical feature extraction from audio. pages 127–130.

Lee, Y. and Fathia, R. N. (2016). Interactive music visualization for music player using processing. In *2016 22nd International Conference on Virtual System & Multimedia (VSMM)*, pages 1–4. IEEE.

Li, T. and Ogihara, M. (2003). Detecting emotion in music.

MakarandVelankar (2020). Mer500.

Margounakis, D. and Politis, D. (2006). Converting images to music using their colour properties. Georgia Institute of Technology.

McEnnis, D., McKay, C., Fujinaga, I., and Depalle, P. (2005). jaudio: An feature extraction library. pages 600–603.

Nguyen, V., Kim, D., Ho, V., and Lim, Y. (2017). A new recognition method for visualizing music emotion. *International Journal of Electrical and Computer Engineering*, 7:1246–1254.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178.

Santos Luiz, C., Mónico, L., Silva, C., and Campelos, S. (2015). Relationship between mathematics and music: Systematization of contents according to the programs of mathematics from 5th to 12th grades of portuguese education.

Soleymani, M., Caro, M., Schmidt, E., Sha, C.-Y., and Yang, y.-h. (2013). 1000 songs for emotional analysis of music. pages 1–6.

Sorussa, K., Choksuriwong, A., and Karnjanadecha, M. (2020a). Emotion classification system for digital music with a cascaded technique. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, 14(1):53–66.

Sorussa, K., Choksuriwong, A., and Karnjanadecha, M. (2020b). Emotion classification system for digital music with a cascaded technique. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, 14:53–66.

Sorussa, K., Choksuriwong, A., and Karnjanadecha, M. (2020c). Emotion classification system for digital music with a cascaded technique. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, 14:53–66.

Ståhl, A., Sundström, P., and Höök, K. (2005). A foundation for emotional expressivity.

Taylor, R. P. (2006). Reduction of physiological stress using fractal art and architecture. *Leonardo*, 39(3):245–251.

Thayer, R. E. (1989). *The biopsychology of mood and arousal*. Oxford University Press Inc.

Weninger, F., Eyben, F., and Schuller, B. (2014). On-line continuous-time music mood regression with deep recurrent neural networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5412–5416.

Whiteford, K. L., Schloss, K. B., Helwig, N. E., and Palmer, S. E. (2018). Color, music, and emotion: Bach to the blues. *i-Perception*, 9(6):2041669518808535.

Yang, Y., Lin, Y., Su, Y., and Chen, H. H. (2008). A regression approach to music emotion recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2):448–457.

Yang, Y.-H. and Chen, H. H. (2012). Machine recognition of music emotion: A review. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 3(3):1–30.

Yang, y.-h., Lin, Y.-C., Su, Y.-F., and Chen, H. (2008a). A regression approach to music emotion recognition. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16:448 – 457.

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

Zhao, S., Gao, Y., Jiang, X., Yao, H., Chua, T.-S., and Sun, X. (2014a). Exploring principles-of-art features for image emotion recognition. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 47–56.

Zhao, S., Yao, H., Wang, F., Jiang, X., and Zhang, W. (2014b). Emotion based image musicalization. In *2014 IEEE International conference on multimedia and expo workshops (ICMEW)*, pages 1–6. IEEE.