Exploring Automatic Translation between Affect Representation Schemes of Music Affective Content

Alissia Ruginǎ[1]

Supervisor(s): Chirag Raman[1]**, Bernd Dudzik**[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 25, 2023

Name of the student: Alissia Ruginǎ
Final project course: CSE3000 Research Project
Thesis committee: Chirag Raman, Bernd Dudzik, Alan Hanjalic

Abstract

Studies in Music Affect Content Analysis use varying emotion schemes to represent the states induced when listening to music. However, there are limited studies that explore the translation between these representation schemes. This paper explores the feasibility of using machine learning models to translate from a dimensional scheme of Valence, Energy and Tension, to a categorical emotion of Anger, Fear, Happy, Sad, or Tender, specifically in the context of musical stimuli. Additionally, this paper considers how the close proximity of certain emotions in the dimensional space, such as Fear and Anger, negatively influence the performance of translation models. This paper reflects on past studies and presents new results concluding feasible translations of music affect content, moreover, providing suggestions for future analysis.

1 Introduction

Affect Content Analysis refers to the study of analyzing emotional states induced using content as stimuli. Content generally explored includes images, videos, text, or audio. This paper will specifically explore the affect states induced by music content as stimuli.

Music is a powerful tool, it has a significant impact on mood regulation and emotional well-being. People may use it to manage their emotions, such as to lift their spirits when feeling down, calm their nerves when feeling anxious, or energize themselves when feeling lethargic. [12] For these reasons there are many studies in the domain of Music Affect Content Analysis that aim to better understand the relationship between music and the emotional states that are induced by listening to it. The analysis of emotion-inducing music can be applied to many fields, including therapy, film-making [2], song recommendations [10], marketing, and to enhance the experience of end-users in any situation where music may be used.

When discussing emotions, there are many different representation schemes used in the field of Psychology to express the multiple emotional states one may feel. Certain representations are Discrete, such as Elkman's Six Basic Emotions where an emotional state is represented as a single categorical emotion: Anger, Fear, Surprise, Happiness, Disgust, or Sadness. There are also Dimensional models that consider emotions as points in a multi-dimensional space, a popular scheme is Russel's Circumplex model, which captures emotion as two-dimensional (Arousal and Valence). [4]

Research in Affect Music Analysis utilizes these different representation schemes to express the emotional states. Though certain schemes are more popular than others, the chosen scehmes differ across studies. The task of translating between representation schemes refers to expressing one emotional scheme in terms of another. For example, given an emotional state described by Valence and Arousal, how may it be expressed as a single categorical emotion? The research of translating between schemes is relevant as there are so many different representations when it comes to emotional states induced by music. To accurately deepen the research of Music Affect Content Analysis, it is important to first understand the relationship between the different representations of emotional states. [6] Studies often use only one representation scheme, consequently it is difficult to compare to other studies that use a different representation. Translation between representations can help bridge the gap between different emotional schemes and provide access to more data that can be combined to facilitate more extensive research.

Problem Statement

This paper explores the task of automatic translation between affect representation schemes of Music Affective Content, through the use of machine learning methods. More specifically, this study answers whether this explored translation task is feasible, and additionally, explores whether there are variables that impact the performance of the translation, such as the close proximity of emotions in the dimensional space.

Relevant research that has influenced this report is discussed in Related Research (Section 2). The process used to answer these questions is explained in Methodology (Section 3), findings are shown in Results (Section 5). The Discussion (Section 6) interprets and analyses the research findings. The Limitations (Section 7) section discusses issues and constraints of the research process. The Conclusions (Section 8) presents final takeaways from the study and what further research may be relevant. Lastly, Responsible Research (Section 9) discusses ethical concerns considered throughout the research study.

2 Relevant Research & Literature

This section presents past studies that are relevant to exploring translation between representations schemes in the context of Music Content as stimuli. These studies presented ideas that were considered for this research study.

2.1 Music as Stimuli

A study by Eerola & Vuoskoski [6] investigates different emotion representations using music as stimuli, however, it does not focus on the translation between schemes, as this paper aims to. Eerola & Vuoskoski's study explores the effectiveness of discrete and dimensional models in the study of perceived emotions in music, as well as determining whether the discrete and dimensional models could be merged or eliminated. The study's results suggest that the three-dimensional model of emotions may be collapsed into a two-dimensional one when applied to music.

Guevara and Eerola's report [5] reasons that music presents affective information of dimensional values such as Valence and Arousal, but not discrete emotions. They argue that the music's attributes may be interpreted as discrete concepts as listeners quickly associate their psychological state with one of the discrete emotions they are constrained to select from. This theory suggests that affect music may be mapped to dimensional representation schemes reliably, however, mappings to discrete emotions (such as the basic emotions) may be imprecise. In regards to the translation task this study explores, this causes reasons to expect inaccuracies when trans-

lating from dimensional to discrete affect representations of musical stimuli.

2.2 Translation between Representation Schemes

A study by Buechel and Hahn, implemented a mapping scheme for Discrete and Dimensional Emotion Representations [3], similar to the goal of this research paper, however, their data was collected from experiments using textual stimuli, rather than music stimuli. Buechel and Hahn's mapping approach relied solely on k-Nearest Neighbours regressions, their study results conclude that significantly accurate mapping may be made back and forth between the discrete and dimensional schemes for textual stimuli, this leads to believe that a mapping for music stimuli may be feasible as well, and suggests the use of k-Nearest Neighbors as a potential translation model.

2.3 Schimmack & Grob model

The three-dimensional emotion model proposed by Schimmack & Grob in their study [16], is based on Russel's Circumplex model [14] of Valence-Arousal, however it proposes to split Arousal into two axes. The proposed three-dimensional model uses Valence, Energy, and Tension axes. Valence refers to a scale of pleasure to displeasure, Energy is described as a scale from awake to tiredness, and Tension scales from tensed to relaxed. A visual representation of the model compared to the Valence-Arousal model is shown in Figure 1. A clear understanding of this model is important as it is used for this study's translation task.

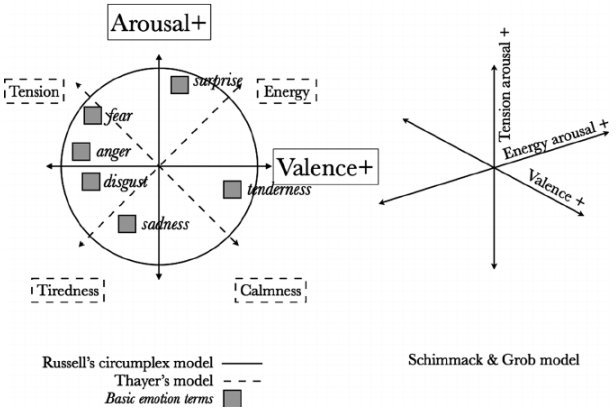

Figure 1: This diagram is reproduced from [6]. It shows a schematic diagram of dimensional models. Note that the axes of the three-dimensional model are not necessarily orthogonal in actual affect data as depicted here.

3 Methodology

Once the topic of Affect Representations and Music Affective Analysis had been researched sufficiently, it was possible to start implementing the research methods in order to explore the feasibility of the translation task. This section will start with an overview of the used datasets, followed by describing the initial analysis, and a description of machine learning models that were considered. The implementation code can be accessed at the github repository https://github.com/alissiarugina/RP_affect_translation/tree/main.

3.1 Datasets

The first step in the research is to identify suitable datasets that may be used for the translation task. A suitable dataset entails a dataset that explores states of Music Affect Content Analysis. These datasets must have annotations that include multiple representation schemes (at least two distinct schemes). The requirement of including more than one type of representation scheme is necessary to perform translation between two different representations, as is the purpose of this project. Finding such datasets was completed by exploring past research in the field to discover accessible and reliable datasets. The suitable dataset that was identified and used throughout this study is: "Film Soundtracks as Stimuli", it is described below.

Film Soundtracks as Stimuli
This dataset collected by Eerola and Vuoskoski [6]comprises 360 excerpts of movie soundtracks that have been annotated by participants. The study uses two different representation schemes, a categorical model and a dimensional model. The schemes are the following emotion models:

- Basic Emotions (Categorical): Happy, Sad, Fear, Anger, Tender. These emotions are derived from the Six Basic Emotions [11] scheme, however, Surprise and Disgust are rarely expressed in music, so Tender was taken instead from the Geneva Emotion Music Scale model (GEMS) [20].
- Schimmack & Grob model (Dimensional): A three-dimensional model that uses Valence, Energy, and Tension. As explained in Section 2.3.

The music selection for stimuli consisted of 360 movie soundtrack excerpts each between 10-30 seconds long, and do not contain lyrics, dialogue, or sound effects. The excerpts were chosen by expert musicologists, each selection aimed to induce a specific emotional state to result in a selection equally representative of the discrete emotion and three-dimensional models.

Each annotation includes scores for the dimensional scheme Valence-Energy-Tension, as well as for each of the following categorical emotions: Anger, Fear, Happy, Sad, and Tender. Scores are given on a scale of 1 to 9. The annotations were completed by 116 university students aged 18–42 years (mean 24.7, SD 3.75, 68% females and 32% males). The resulting ratings have been averaged over multiple participants.

3.2 Dataset Manipulation & Analysis

Dataset manipulation was required to alter the collected dataset *Film Soundtracks as Stimuli* to a format that is suitable for the Machine Learning task and will ensure the intended conclusion can be drawn correctly.

Since categorical representations schemes are generally taken as a single emotion, this dataset was adapted to add a new column "Max_emotion" that takes the categorical emotion with the highest score out of the annotated scores of

[Anger, Fear, Happy, Sad, Tender]. This would allow translation mappings from the [Valence, Energy, Tension] scores to a single [Max_emotion] category, thus a Classification problem. Data analysis was done to ensure validity and reliability. It was checked for repeated entries, outliers, null entries.

Data Analysis

It was important to first analyze the data set and its properties to ensure that suitable machine learning models are selected, as models often make assumptions about the data it fits.

The class distribution was analyzed, where class refers to the categorical emotions represented by the values in "Max_emotion" column. The frequency of each emotion class is shown in Figure 2. It is evident that the classes are not of equal frequencies, there are noticeably fewer data points classified as "Happy" or "Anger", this must be considered when training a machine learning model.

Data Visualisation is a useful tool to analyze the data, it may help identify outliers, trends, and patterns in the data [18]. To visualize the data, a 3D scatter plot of the *Film Soundtracks as Stimuli* data set was analyzed and is shown in Figure 3. This plot alone cannot draw any particular conclusions, though the distinct colors that represent the categorical emotions do appear to be clustered by their color groups, seems likely that there is a significant relationship between the location on the Valence, Energy, Tension axis' and the corresponding discrete emotion. The averages and standard deviations of the Valence, Energy, and Tension scores were analyzed per categorical emotion class, shown in Table 1. Analyzing the class means, Fear and Anger's Valence, Energy, and Tension means are quite close in value, however this must be further analysed to draw any conclusions.

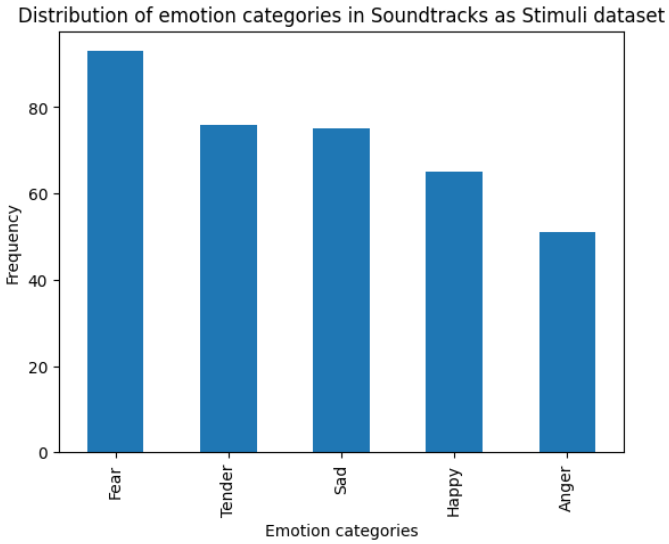

Figure 2: Bar graph displaying the class distribution of the Soundtracks as Stimuli dataset, wehre each bar shows the frequency of the corresponding emotion class.

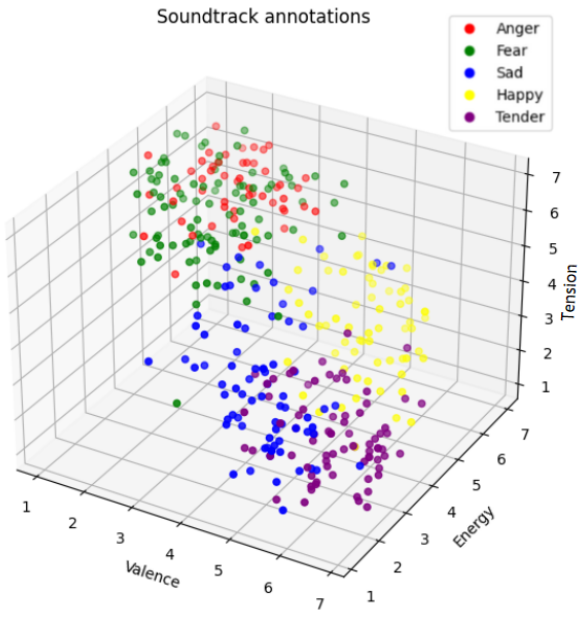

Figure 3: Soundtracks data points plotted on Valence, Energy, Tension axes. Colored according to the corresponding emotion category.

Table 1: Average Valence-Energy-Tension Values per emotion class

Emotions	Valence		Energy		Tension	
	avg	sd	avg	sd	avg	sd
Anger	2.397	0.671	5.387	0.820	6.107	0.462
Fear	2.434	0.851	4.620	1.067	5.920	0.650
Sad	4.600	0.809	2.656	1.001	3.429	1.099
Happy	5.206	0.755	5.213	1.002	3.458	0.816
Tender	5.871	0.629	2.925	0.717	2.247	0.764

3.3 Suitable Machine Learning Models

The translation task that was explored was translating from Dimensional Emotion Schemes to Categorical Emotion Schemes. Using the *Film Soundtracks as Stimuli*, the Dimensional Scheme, in this case, is [Valence, Energy, Tension], and the Categorical Emotion Scheme is the corresponding single categorical emotion from the list [Anger, Fear, Happy, Sad, Tender] (called "Max_emotion" in the dataframe). Using the dimensions of Valence, Energy, and Tension to map to a category of emotions is a multi-class classification problem. Three distinct supervised machine learning models were explored to implement this translation, a linear classifier: Logistic Regression Classifier, and two non-linear models: Decision Tree Classifier, and K-Nearest Neighbors Classifier.

Logistic Regression

A Logistic Regression uses a logistic function to model the relationship between the independent variables and the dependent variable. It is generally used for predicting binary

target variables, however, a variation of a Logistic Regression, the Multinomial Logistic Regression, may be applied for multi-class classification [13].

Applying the Logistic Regression to the Categorical Emotions classification problem, the model uses the dimensional scores of Valence, Energy and Tension, as predictor values to map to each of the 5 categorical emotions (Anger, Fear, Happy, Sad, or Tender).

Decision Tree Classifier

A Decision Tree Classifier is a supervised machine learning model that uses a tree-like model to make decisions, or splits, based on features in the input data and specified splitting criteria. Each internal node represents a decision, each branch is an outcome of that decision, and each resulting leaf node refers to a class label [7].

For this translation task, the Tree Classifier makes splits based on the dimensional scores of Valence, Energy and Tension, until reaching a final categorical emotion (Anger, Fear, Happy, Sad, or Tender). A Decision Tree model can be of any number of levels and may use various splitting criteria such as "gini", "entropy", "log loss".

K-Nearest Neighbors Classifier

The K-Nearest Neighbors Classifier is based on the assumption that similar things exist in close proximity. It predicts the class of a new data point by finding the 'k' nearest data points in the training data and predicts the new point's class to match the mode of the k-nearest points' classes. It is non-parametric, thus makes no assumptions about the data's distribution. [8]

4 Experimental Setup

This section describes the exact steps taken to implement the exploration task of translation. Firstly, the chosen evaluation methods are explained, followed by the optimization techniques used to select parameter values for each of the chosen models.

4.1 Evaluation Method

The metric of accuracy was chosen to evaluate the machine learning models' performance of the translation task. Accuracy is calculated as the fraction of correctly classified samples, predicted by a model given the Valence, Energy and Tension values.

To ensure reliable performance values, a 5-fold cross-validation method, repeated for 20 different fold splits, was used such that the compared metrics are averaged over multiple folds and repetitions and not dependent on a specific split. The cross-validation technique was performed by splitting the data into 5 folds using Scikit Learn's StratifiedKFold [1] function which generates folds that preserve the percentage of samples for each class, this is used to account for the class imbalances in the dataset. Then, evaluating the performance by taking 4 of the folds as training data and measuring the performance on the one remaining validation fold, this is done 5 times, with each fold taken as a validation set once. This 5-fold evaluation was repeated 20 times using different Stratified Folds each time. The accuracy metric taken as representative of a model's performance is the average of the validation accuracy scores from the 20 repetitions of 5-folds, thus the average of 100 validation scores.

Baseline Model

A Dummy Classifier[2] , from Scikit Learn's library, was used as a Baseline Model using the 'most_frequent' strategy. This strategy constantly predicts the most frequent class label in the y arguments used to fit the model, thus performs as good as random chance when class distributions are equal. Using it as a baseline comparison for translation models indicates that a given model is better than random chance if its performance is significantly different from the dummy model's performance.

4.2 Decision Tree Optimization

Using Scikit Learn's library for Decision Trees[3] , one of the first parameters to select is the splitting criteria, the options are "gini", "entropy", "log loss". Gini was selected as it is particularly useful for multi-class classification problems, thus it is suitable for this task.

The depth of the Decision Tree was optimized by comparing the accuracy outcomes of trees with varying depths, using repeated k-fold cross validation [1]. The performance measure compared was accuracy over 5-fold cross-validation, repeated using 20 different splits, where each decision tree model used the same splits to keep it constant. The average accuracies for each depth value is plotted for both the training sets and the validation sets, as shown in Figure 4. By comparing the performance on the training set vs. the validation set, the models with depths greater than four are likely overfitting the training data as the performance on the training set increases though the performance on the test data decreases. For this reason, a depth of 3 was chosen as optimal.

[1] $https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.StratifiedKFold.html$

[2] https://scikit-learn.org/stable/modules/generated/sklearn.dummy.DummyClassifier.html

[3] https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeClassifier.html

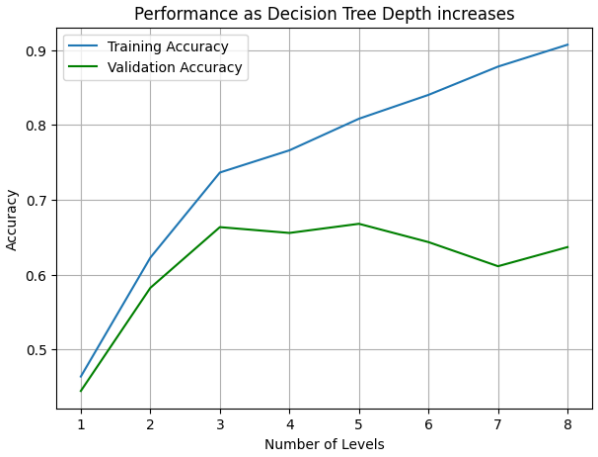

Figure 4: Graph of Decision Tree Classifier's accuracy on Training and Test sets, using increasing depths. Used to observe optimal depth of model.

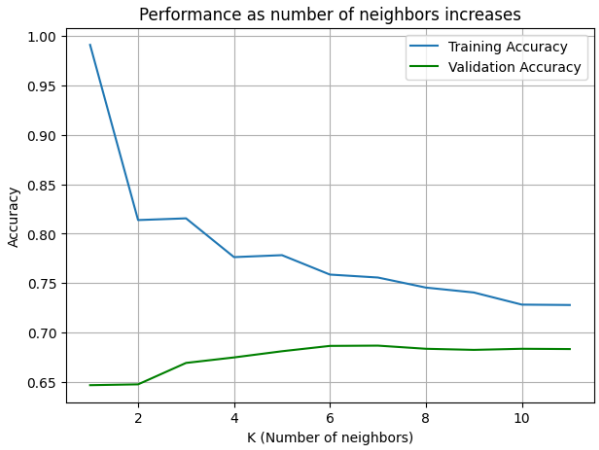

Figure 5: Graph of K-Nearest Neighbors' accuracy on Training and Test sets, using increasing K values. Used to observe optimal K value.

4.3 Logistic Regression Optimization

When using Scikit Learn's library for Logistic Regression[4], there are multiple parameters that may be selected according to the data assumptions. For the translation task, the multinomial condition was selected, since the task deals with multiclass classification. The maximum number of iterations parameter ("max_iter") was adjusted since using the default value (100) consistently lead to errors due to max iterations reached. Increasing values were tried until a value of 600 generally terminated without reaching the max iterations, thus 600 was chosen as the parameter's value. Another parameter, "C", the inverse of regularization strength can take any positive float value. The performance of the training and test sets using different "C" values were plotted to decide on an optimal value, however, the different values resulted in negligible differences in performance, thus no optimal "C" value may be concluded and so the default value of 1 was taken.

4.4 K-Nearest Neighbors Optimization

The Nearest Neighbors was implemented using Scikit-Learn's KNeighborsClassifier[5] The performance of the K-Nearest Neighbors algorithm differs depending on the chosen K value. To compare the performances of using different K values, the training and test accuracy was graphed on increasing K values, as shown in Figure 5. Based on the graph, a K value of 6 was chosen since it achieves the peak accuracy score for the validation set.

Another parameter to consider is the distance metric used when calculating the distance between points. Commonly used distance measures include Euclidean, Manhattan, and Minkowsky Distance. Minkowsky was used since it is a generalized form of Euclidean and Manhattan distance metrics.

[4]https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html#sklearn.linear_model.LogisticRegression

[5]https://scikit-learn.org/stable/modules/generated/sklearn.neighbors.KNeighborsClassifier.html

5 Results

The results of the translation task are described in this section. This entails the different classifiers' performances, evaluated as explained in the previous section, as well as a description of the models' performances across the specific emotion classes.

5.1 Classifiers' performance

Table 2 displays the overall performances of the optimized Decision Tree Classifier, Logistic Regression Classifier, and K-Nearest Neighbors Classifier (KNN). The performance metric I used is the average, minimum, and maximum accuracy scores obtained when using the 5-fold cross-validation method, repeated 20 times. The table also includes comparisons to the Baseline Dummy Classifier's performance when using identical test and validation splits. The table includes the difference in means, this is calculated as the mean accuracy score of the Baseline Classifier subtracted from the mean accuracy score obtained by the chosen given model. The "SD" column represents the standard deviation of the accuracy scores obtained by the model compared to scores obtained by the Baseline Model. The "t-test" shows the t-statistic obtained by a 5-fold paired t-test procedure to compare the performance of the model to the Baseline model, the "p-value" indicates the probability of obtaining the observed difference or a more extreme difference if the null hypothesis is true. Since the p-values are extremely small for all three models, the t-test null hypothesis is rejected at any reasonable significance level, thus suggesting that there is a significant difference between the performance of these models compared to the Baseline model.

Table 2: Performance Metrics of models and comparisons to Baseline Dummy model.

Models	Accuracy			Compared to Baseline			
	average	min	max	means diff	sd	t-test	p-value
Decision Tree	0.666	0.625	0.694	0.408	0.289	26.07	9.56e-05
Logistic Reg	0.700	0.611	0.750	0.441	0.312	25.11	7.64e-05
KNN	0.669	0.611	0.736	0.411	0.290	23.30	6.21e-05

To determine which classifier is best for the translation task, further analysis was done. A One-Way ANOVA test was used to test whether there is a significant difference in performance across the three models. The ANOVA tests for differences in the means of the groups using a variance, the null hypothesis assumes no significant difference across the means of the groups, this hypotheiss is rejected if the p-value is lower than a significant level [17]. Scipy Stat's f_oneway[6] function was used, inputting the three models accuracy measures as samples. The test results are shown in Table 3

Table 3: One-way ANOVA test results comparing performance across the three models

F-value	0.05379
p-value	0.94785

The p-value is greater than any reasonable significance level, thus the test does not observe a significant difference across the performances of the Logistic Regressor, Decision Tree Classifier, and the K-Nearest Neighbors Classifier.

5.2 Performance analysis per class

The accuracy of classifying the correct emotion was observed per each class to analyze whether the model performs equally well across all five emotion labels. Accuracy for this analysis was done using the same explained 5-fold (20 times repeated) cross-validation.

Figure 6 shows the performance of each model, using a radar map to compare the difference in accuracy per class visually. The lowest accuracy is obtained by the Logistic Regressor for the Anger class. Overall, it is shown that all models have the highest accuracy in predicting data points in the Happy class and the lowest accuracy for the Fear and Anger classes.

Having visually observed a difference in model performance per class, I tested the significance of these differences using three ANOVA tests, one for each model, testing the model's accuracies using the five emotion classes as separate groups. All ANOVA tests resulted in a p-value less than 0.05, thus concluding a significant difference between the accuracies over each class.

The models' performances are lowest when classifying Anger and Fear annotations. I explored the possible reasoning for these results. All three classifiers that were attempted for the translation task relied on the annotations' features of Valence, Energy, and Tension scores to distinguish between the different discrete emotion classes. The differences in these classes' features was verified and proven by an ANOVA test. Afterwards, a Tukey pairwise comparison, using statsmodels' pairwise_tukeyhsd[7] was used to determine whether the classes' features were significantly different for each of the dimension features, the adjusted p-values (adjusted to account for multiple comparisons) are shown by the heatmap in Figure 7. Tukey's HSD test is a post-hoc test often used after conducting an ANOVA to compare the all the pairwise group differences while controlling for family-wise error rate. The test assumes the same hypotheses as the ANOVA test (null hypothesis: the means are all equal, alternate: the means are significantly different) [19]. The Tukey results indicate that the annotations that classify as Anger and Fear do not differ significantly in Valence and Tension scores (as the p-value is greater than 0.05 thus the null hypothesis cannot be rejected), however, they do differ in Energy scores. This similarity in two out of three of the features may be the

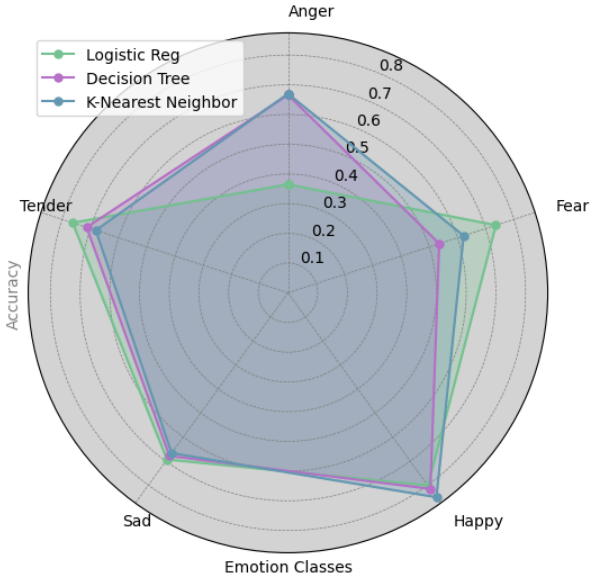

Figure 6: Radar map of the average accuracy performances per emotion class, each model shown by the colour specified in the legend.

[6] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.f_oneway.html

[7] https://www.statsmodels.org/dev/generated/statsmodels.stats.multicomp.pairwise_tukeyhsd.html

reason for the models' difficulties in distinguishing between Fear and Anger in the dimensional emotion space.

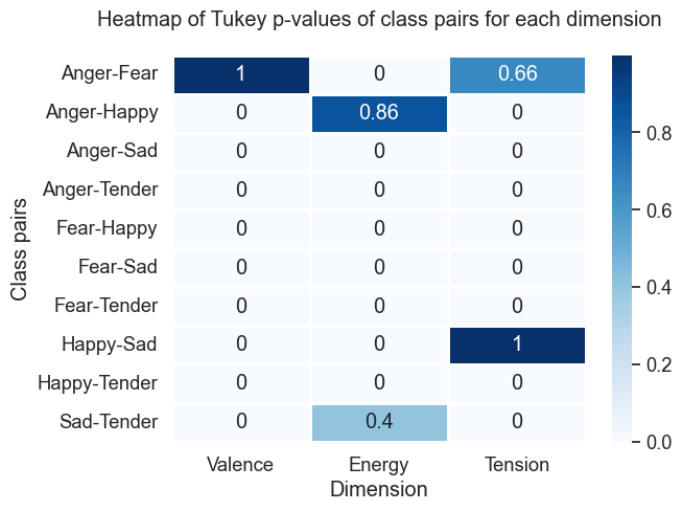

Figure 7: Heatmap that displays the adjusted p-value obtained using the Tukey Pairwise Honest Significance Test comparing the significant difference in each dimension (Valence, Energy, and Tension) between each pair of emotion classes.

6 Discussion

The main aim of this study was to explore the feasibility of a translation from dimensional emotion representation schemes to discrete representations in the context of musical stimuli. It should be noted that this study solely explores the specific translation of using Schimmack & Grob's model of Valence, Energy, and Tension scores to map to a discrete emotion of the following categories: Anger, Fear, Happy, Sad, Tender. Additionally, the translation has only been evaluated on data from the same dataset the models were trained on, meaning it is unknown how this model would generalize to data that was collected from a different study.

By evaluating the performance of three different machine learning classifiers on the Soundtracks as stimuli dataset, and comparing their accuracy to a Baseline Dummy Classifier, we can conclude this specific translation task as feasible since all models performed significantly better than the Baseline.

I explored three different classifiers: Logistic Regression Classifier, a Decision Tree Classifier, and a K-Nearest Neighbor Classifier, to explore which model would fit the translation task best. After comparing the models' accuracies (Section 5.1) and finding no significant differences in performance across models (ANOVA test in Table 3), I cannot conclude whether one of the three models performs better than the others, considering only accuracy.

The secondary research aim of this study was to analyze which factors of the dataset impact the performance of the translation. I observed that the models' performed significantly different across the different discrete emotions. More specifically, they all had lowest accuracy results in classifying Anger and Fear representations. By analyzing the pairwise class differences in Valence, Energy, and Tension scores (Tukeys pairwise hsd results shown in Figure 7) I found that Anger and Fear representations differ significantly in only Energy scores, not Valence or Tension scores. This similarity in two out of three of the features, and consequently their close proximity in the dimensional space may be the reason for the models' difficulties in distinguishing between Fear and Anger in the dimensional emotion space. A previous study by Scherer [15] presents a similar idea that music as stimuli can only induce a restricted range of unpleasantness and activation, in consequence, rating listeners' states using Valence and Activation may not allow for a strong separation of emotions. The study concludes that describing the emotional states induced by music using valence-activation approaches is more realistic than using categorical emotion approaches, the same theory is confirmed by Guevera's study (previously discussed in Section 2). Reflecting upon these findings, perhaps using categorical schemes to describe music induced states is not the most accurate representation, and so the translation is not all that meaningful or accurate.

Future translation studies using Music Affect Content Analysis should select categorical emotions that are known to be more widely dispersed in the dimensional emotion space. Alternatively, the distinction of Fear and Anger (induced by music) in the dimensional space can be studied further by exploring whether using information about the music excerpt can help differentiate between the two, as the study [9] investigates the musical properties that differ between discrete emotions, including the differences between Fear and Anger in acoustic properties such as sound level and timbres. A future study should investigate if the use of acoustic properties can be applied to improve translation between representations.

7 Limitations

Throughout my completion of this research project, some factors constrained the depth of the study, such as learning the interdisciplinary topics, finding datasets, and the time constraint of the project.

The interdisciplinary aspect of the study meant that I first had to spend time on simply understanding the research topic of Affect Representation Schemes and Content Analysis, as I had limited knowledge about this field previous to the start of the project.

Finding suitable datasets to be used for the study was a difficult and timely task. A suitable dataset had to meet specific requirements, thus I spent the first few weeks of the project searching for datasets. I found only two suitable datasets which did not complement each other, consequently, the dataset that was chosen, *Film Soundtracks as Stimuli* had limited attributes included. The dataset had no specific information about the annotations' participants, nor any attributes of the song excerpts (besides a Soundtrack name and excerpt duration). The limited information led to a general translation task and limited analysis. Given a dataset with more variables, a more interesting and detailed analysis may have been drawn.

The limited time given for the completion of this project certainly led to a less extensive study. Certain decisions as

well as analyses had to be rushed at times to ensure project completion by the given deadline.

8 Conclusions and Future Work

In this study I answered the research questions; exploring whether a translation between representation schemes induced by music stimuli is a feasible task, and additionally, exploring what factors may influence this translation. After researching past studies and analyzing the dataset I found, I modeled the translation from Schimmack & Grob's dimensional model of Valence, Energy, and Tension to categorical emotions of Anger, Fear, Happy, Sad, and Tender, using three different models; a Logistic Regressor Classifier, a Decision Tree Classifier, and a K-Nearest Neighbors Classifier. I optimized their parameters, then assessed their performance using accuracy and comparing to a Baseline Dummy Classifier. All three models performed significantly better than the baseline model, thus implying the translation task as feasible.

Comparing the models' performances across the five different discrete emotion classes, I found that models struggled to differentiate between Fear and Anger annotations in the dimensional space. The proximity of Anger and Fear in the Valence and Tension axes may be causing the difficulty in differentiating between the two classes, and consequently, negatively influencing the translation accuracy. Future studies that explore translation tasks of Music Affect Content should consider using discrete emotions that are more dispersed in the dimensional space. Otherwise, the distinction of Fear and Anger in the dimensional space can be studied further by exploring the use of acoustic features in the translation.

9 Responsible Research

As with any research project, it is important to consider the ethical concerns of this study and to conduct the necessary research responsibly. This section discusses the approach taken to ensure responsible data collection, and additionally, presents the ethical implications this study's findings may induce.

9.1 Data collection

The research methods were performed responsibly and ethically. No raw data was directly collected for this research study. The dataset used for analysis was found online, consists of results from a past experiment conducted by experts in the field, involving human participants. The original collection of data adhered to the ethical guidelines of experiments with human subjects, and obtained consent from participants. The dataset source is cited accordingly. The reports describing the data sets were read carefully, and no further assumptions were made about the data. No changes have been made to alter the data results or findings, the findings and conclusions drawn from this dataset are based on the careful analysis and interpretation of the existing data.

9.2 Ethical use of findings

This study intends to achieve a model that accurately translates between representation schemes of Music Affect Content. This translation task on its own has no ethical implications, however, the research it encourages may have ethical concerns to be considered. The translation task promotes Affective Computing, and consequently, its applications in the real world. This refers to using music to induce people to feel specific emotional states, certain applications of this research may be considered a form of manipulation or may lead to exploitation

References

[1] Rohit Bhadauriya. Cross-validation(cv) and hyperparameter tuning, Sep 2021.

[2] Fernando Bravo. The influence of music on the emotional interpretation of visual contexts. volume 7900, 01 2011.

[3] Sven Buechel and Udo Hahn. A flexible mapping scheme for discrete and dimensional emotion representations: Evidence from textual stimuli.

[4] Paul Buitelaar, Ian D. Wood, Sapna Negi, Mihael Arcan, John P. McCrae, Andrejs Abele, Cécile Robin, Vladimir Andryushechkin, Housam Ziad, Hesam Sagha, Maximilian Schmitt, Björn W. Schuller, J. Fernando Sánchez-Rada, Carlos A. Iglesias, Carlos Navarro, Andreas Giefer, Nicolaus Heise, Vincenzo Masucci, Francesco A. Danza, Ciro Caterino, Pavel Smrž, Michal Hradis, Filip Povolný, Marek Klimeś, Pavel Matějka, and Giovanni Tummarello. Mixedemotions: An open-source toolbox for multimodal emotion analysis. *IEEE Transactions on Multimedia*, 20:2454–2465, 9 2018.

[5] Julian Cespedes-Guevara and Tuomas Eerola. Music communicates affects, not basic emotions - a constructionist account of attribution of emotional meanings to music. *Frontiers in Psychology*, 9, 2 2018.

[6] Tuomas Eerola and Jonna K. Vuoskoski. A comparison of the discrete and dimensional models of emotion in music. *Psychology of Music*, 39:18–49, 2011.

[7] Prashant Gupta. Decision trees in machine learning, Nov 2017.

[8] Onel Harrison. Machine learning basics with the k-nearest neighbors algorithm, Jul 2019.

[9] Patrik N. Juslin and Petri Laukka. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*, 129:770–814, 9 2003.

[10] Stuart Ough Chin-Chang Ho Karl F. MacDorman. Automatic emotion prediction of song excerpts: Index construction, algorithm design, and empirical comparison. *Journal of New Music Research*, 36(4):281–299, 2007.

[11] MSEd Kendra Cherry. The 6 types of basic emotions and their effect on human behavior, Dec 2022.

[12] Adam J. Lonsdale and Adrian C. North. Why do we listen to music? a uses and gratifications analysis. *British Journal of Psychology*, 102:108–134, 2011.

[13] Ashwin Raj. Perfect recipe for classification using logistic regression. *Towards Data Science*, Jan 2021.

[14] James Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39:1161–1178, 12 1980.

[15] Klaus R. Scherer. Which emotions can be induced by music? what are the underlying mechanisms? and how can we measure them? *Journal of New Music Research*, 33:239–251, 2004.

[16] Ulrich Schimmack and Alexander Grob. Dimensional models of core affect: A quantitative comparison by means of structural equation modeling. *European Journal of Personality*, 14:325–345, 2000.

[17] Julia Simkus. What is an anova test in statistics: Analysis of variance, Jun 2023.

[18] Gilbert Tanner. Introduction to data visualization in python. *Gilbert Tanner*, Jan 2019.

[19] Zach. A guide to using post hoc tests with anova, Apr 2019.

[20] Marcel Zentner, Didier Grandjean, and Klaus R. Scherer. Emotions evoked by the sound of music: Characterization, classification, and measurement. *Emotion*, 8:494–521, 8 2008.