Affect Representation Schemes used in Music Automatic Affect Prediction

A Systematic Literature Review

Natalia Bryła[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: Natalia Bryła
Contact: n.m.bryla@student.tudelft.nl
Final project course: CSE3000 Research Project
Thesis committee: Chirag Raman, Bernd Dudzik, Cynthia Liem

Abstract

There is a correlation between music and affect which researchers try to define and use in technology to improve healthcare and users' experience in music-related technology. However, since affect is a complex term there is not a specified way on how to represent different affective states in systems. A systematic literature review was performed to give an overview of affect representation schemes (ARS) used in music affect prediction. All the studies included in this survey were found in *Scopus, Web of Science* and *IEEE Explorer* and focus on automatic affect prediction in music. What is more, the literature written in a different language than English and not from the Computer Science field was excluded. Considering the time constraint of 10 weeks, an additional feasibility filtering was conducted. That is only the documents which mention common benchmark datasets for this task were included. The datasets were chosen after the research of related work and different literature reviews related to this topic. We were aware that it might include a bias correlated to dataset popularity. At the final stage of filtering, 113 records were chosen for data extraction from which 51 were used in the analysis due to time constraints and a single person conducting this research. After synthesising and analysing the extracted data we observed that the schemes can be distinguished between *dimensional* and *categorical* approaches and both are similarly popular. A rare but existing ARS are developed by combining both scheme types in a unique representation. What is more, there is no easily visible trend over time in the usage of certain schemes and only 66% of the reviewed studies support their choices with psychological theories. While conducting this research, we encountered limitations in terms of time and the research group, so we leave a gap for future improvements in this project.

1 Introduction

Emotions play a significant role in people's interactions both with other humans and computers. Even though they accompany people every day [1] defining emotions is a complex problem [2]. What is more, people tend to confuse emotions with affects, where the affect refers to various mental states, which can be emotions, feelings, moods or attitudes [2]. On the other hand, emotions are triggered by external stimuli. Even though there is a study that analyses the differences between emotions and those affects [2] some researchers cover a variety of emotional states under the name of affect [3, 4].

In the past years people started to be more interested in benefiting from combining emotional affects understanding with technology and one interesting part of it is music emotion recognition (MER). It is a fascinating problem since music is constantly present in life and can induce different emotions in listeners [5]. Solutions to MER are usually based on automatic affective computing, which is the study and development of systems that were made to recognize, express, and have emotions [6]. Automatic affect prediction aims to use computational methods like machine learning or deep learning algorithms to analyze input data and predict people's invoked affects or music affective content. The music affective content is the features of music that can be analysed and used for computations to predict the emotions that people perceive from this audio. Additionally, according to [7], there exists a correlation between music and the way people react to it and because of that the affect prediction technology might be used to help specialists improve music therapy [8] and better understand mental health areas [9]. What is more, precise affects prediction could be used to enrich customers' and users' experience while using everyday technology.

There is already literature about approaches to resolving MER problems [10, 11], however, the representation of emotions and affective states is not unified. The way in which the affective states are represented, or how their affect representation schemes (ARS) look, is a crucial aspect of affective computing. The ARS are used to represent and differentiate between affects for computational modelling. However, there is no common way chosen by the experts on the best approach to do the representation. This is because in practice different representations suit best for different needs [1]. As a result, researchers often employ a variety of representation schemes, sometimes creating novel ones not based on existing psychological theories.

MER was already analysed by researchers, however, they mainly focused on different machine learning approaches and the use cases of those techniques [12, 13]. There is also research on common emotion models used in computer science [1]. Nonetheless, there is no work describing the affective representation schemes used in music automatic affect prediction or MER.

Having this background knowledge this survey was conducted to fill the knowledge gap and answer the research question *RQ1*:

Which Affect Representation schemes are used in Automatic Music Affect Prediction Systems ?

To provide a detailed and precise overview of current work the following 7 sub-questions were answered:

1. *(SQ1) What types of affective states have been targeted by prediction systems?* Targeted affective states give a meaningful insight into the connection between different affects and the optimal fit for the representation scheme. What is more, identifying the target and analysing the system might show the confusion between the understanding of affective states and emotions.
2. *(SQ2) What different affect representation schemes have been used for this, and if so, what is the motivation for this particular scheme?* This is a crucial question and answers to it will give the base information about commonly used schemes. That means the data gathered for the purpose of this question would already partially answer the main research question.
3. *(SQ3) Are systems using more than one emotion representation scheme simultaneously, and if so, what is their motivation for doing so?* During reading the literature we might encounter several different approaches to affect representation. An interesting idea could be the combination of already existing and used schemes to obtain greater advantages. Those ideas might be unique and lead to faster growth in this field.
4. *(SQ4) What different types of input data do prediction systems use for their analysis?* Knowledge about the input data to different systems might have an impact on the chosen ARS. Within music it can be expected to use an audio sample as the input, however, there may be various cases where the study focuses on human responses and analyze physiological signals.
5. *(SQ5) Are there differences in the popularity of schemes used for modelling different affective states?* Understanding the popularity of certain affective representation schemes can contribute to a deeper understanding of their profitability, as trends in popularity are often rooted in specific underlying reasons. However, it might be possible that the popularity of certain datasets alters the results of a general trend. Identifying the popular schemes can encourage researchers to use the same representation scheme and enable comparisons between systems.
6. *(SQ6) Has the popularity of specific schemes changed over time?* Learning from existing research speeds up the development and progress of new research. As a consequence, analyzing changing trends over the years can provide valuable insights into significant changes and serve as inspiration for future researchers to pursue specific directions.
7. *(SQ7) Is the majority of representation schemes used based on psychological theory?* Computer scientists might not always focus on psychology theories and augmentations. However, in terms of affects, psychological knowledge can appear to be crucial to reach real-world representation.

A systematic literature review was conducted and it presents an analysis of the current state of the literature on different ARS. This research can be helpful to get a short and descriptive overview of this research field with an in depth analysis of the usage of ARS. A reliable overview of this topic might be used by future researchers to get inspired and build on existing studies. Moreover, it can speed up the process of decision-making of the best matching representation to scientists' needs and accelerate the growth of research in this field.

The remainder of this report is organised as follows. Section 2 describes in detail the methods used during the research process. This section is divided into subsections to separately describe the search strategy, selection process and information extraction parts. Section 3 provides the analysis and synthesis of found data along with the answers to the research question and sub-questions. Next section 4 describes the ethical aspects of the research and discusses the risks and bias. Section 5 presents the general interpretation of the results and discusses any limitations of the review. Finally, section 6 concludes the presented study and shows recommendations for future work.

2 Methodology

This research proposes a systematic literature review of different approaches to representing the affective states in automatic affect prediction systems. We decided to follow the PRISMA [14] checklist to deliver a review in a structured way. The details of the methodology used to perform this review along with the descriptions of certain steps are presented in this section. In subsection 2.1 all eligibility criteria for choosing a relevant set of papers are described. Next subsection 2.2 presents the used search engines and subsection 2.3 shows which search strategy and why was utilized. The brief description of the selection process is presented in subsection 2.4 and search results in subsection 2.5. Finally subsections 2.6 and 2.7 describes the process of data extraction and synthesis correspondingly.

2.1 Eligibility Criteria

Several inclusion and exclusion criteria were defined to get a final set of papers. The final set should include all the relevant papers for this research and be large enough to provide a meaningful overview. However, it should also be manageable in size to analyze the literature during the time of the research process. The table with chosen criteria is presented in a table 1.

Table 1: Inclusion and exclusion criteria.

Inclusion	**Exclusion**
People listening to music in various situations	
English language papers	Papers written in different languages than English
Papers from Computer Science field	Papers from a field other than Computer Science
Research whose main focus is on music automatic affect (or emotion/ mood/ feeling) prediction (or recognition/ detection)	Research that does no automatic computations on music or other input signals
Only journals, conference papers or books chapters	Literature reviews

We decided on these criteria to get an appropriate set of final papers from which we could extract the data needed to answer the research question. We include both the studies that work with people listening to music to predict the affect and the studies that use only music as the input to find the affective characteristics, as those searches should provide us with the relevant literature. In addition, to make this survey feasible and stick to the Computer Science field we excluded all the literature from other fields and written in different language than English.

2.2 Search Engines

To find relevant literature for the survey different databases were considered. Three of all accessible research tools were included for enquiring. One of the chosen databases is *Scopus* [1], it is large in size database with a broad scope and sufficiently high quality. Another included research tool is *Web of Science*[2] a platform which similarly to *Scopus* has access to a very big amount of data of scientific content with broad scope. Lastly, we decided to use *IEEE Explore*[3] since it is a database which provides the users with scientific literature related to computer science, electrical engineering and electronics.

In the first stage of research, several different research tools were under consideration, however, they were not used for the final search. *Google Scholar*[4] was excluded because the amount of obtained results after first filtering was unfeasible to manually screen by one person. Moreover, *ISMIR (The International Society for Music Information Retrieval)* explorer also had to be excluded since the interface is not convenient for systematic searches. It does not support advance search, filters or options to export the list of documents, which would make the process of systematic literature review more complex and time-wise unattainable. Other research tools were not included because of their accessibility, scope, size or searching and exporting options.

[1] https://www.scopus.com
[2] https://www.webofknowledge.com
[3] https://ieeexplore.ieee.org
[4] https://scholar.google.com

2.3 Search Strategy

The research question and sub-questions were already identified by the supervisors and because of that the first step of the systematic literature review was to determine the search terms. The goal was to find the keywords that cover most of the fields that we wanted to investigate. To do this the focus was first put on defining the core concepts. Those concepts were derived from the research questions in a way to describe each part of the research with several synonyms. The core concepts with corresponding search terms are presented in the table 2.

Table 2: Concepts and synonyms used as search terms. The asterisk represents the possible various number of characters to add in its place, for example for emotion* the found words would also be emotional, emotions, emotionally.

Concept	Search Terms
Affect	affect representation schemes, affect representation, emotional states, represent* of affects, feeling*, happy, sad, arousal, angry, anger, fear, surprise*, happiness, emotion*, disgust
Automatic Affect Prediction	automatic affect* prediction, affect* prediction, predict*, emotion recognition, recognition
Music related	music*, audience, listen* to music, songs

Those groups of keywords were joined together to cover the full field necessary for this research. Within those, we defined the search query that was later used for searching the databases and finding relevant papers for this survey. The full search query that was used in this process can be found in appendix A.

Feasibility Filtering

This survey had to be done within 10 weeks and because of that an additional feasibility filtering was performed. To limit the number of obtained result papers before manual filtering we decided to consider only the literature in which some of the known benchmark datasets for music affect recognition were used. The databases were chosen after reviewing other literature surveys regarding similar music-related topics, [12, 13, 15] and the full list of included datasets can be found in the table 3. Additionally, the information on the annotation style and publication year can be also found in this table.

Modification of the search query with those databases enabled to systematically lower the amount of resulted literature. We were aware that this filtering might introduce a bias towards certain ARS. However, those datasets are diverse in types of annotation which lowers the probability of altered results.

2.4 Selection Process

To obtain the final set of papers included in the review the proposed selection process was done. First, we reviewed all the chosen databases, described in 2.2, and tried different queries created based on identified keywords, presented in 2.3. During the screening based on title, abstract and keywords with the use of inclusion and exclusion criteria some papers were included for the later investigation and some were not. If during the process of searching through databases, we realised that the inclusion and exclusion criteria or keywords for the query had to be adjusted the whole process of constructing the search query, searching across databases and selecting final papers was repeated.

When we found the best-performing query and collected the papers from each of the chosen databases we started to systematically exclude irrelevant documents. The filters supported by databases were used correspondingly to the set inclusion and exclusion criteria described in 2.1. Next, all the duplicates given by databases were removed and we followed with manual filtering.

Filtering was performed three times before the final set of papers was defined. First, we filtered the literature base on the title and excluded all the documents with titles not corresponding to the search scope. When filtering by title was finished we followed with filtering by abstract. Since abstracts give a general overview of a content it was possible to exclude more irrelevant papers. The final part of this process was the full-text filtering which was performed in parallel with information extraction that is described in more detail in section 2.6. The selection process of the research was crucial to get a meaningful set with a reasonable size which contains as much relevant literature as possible.

Table 3: Datasets used for feasible filtering with the corresponding annotation and data type.

Dataset	Annotation	Type of data	Year	Ref
CAL500	18 emotions	songs	2008	[16]
MagnaTagATune	Tags	music clips	2009	[17]
DEAP	Valence, arousal and dominance, physiological signals	music videos	2011	[18]
MOODetector:Bi-Modal	Valence and arousal	music excerpts	2011	[19]
MOODetector:Multi-Modal	Mood/color-perception	music excerpts	2011	[20]
MSD	tags	music excerpts	2011	[21]
Soundtracks	Valence and mood	music excerpts	2011	[22]
MoodSwings	Valence and arousal	song clips	2012	[23]
DEAM -the MediaEval Database for Emotional Analysis of Music (mediaeval2013, mediaeval2014, mediaeval2015)	Valence and arousal	excerpts and full songs	2013	[24]
Emusic	Valence and arousal	experimental music excerpts	2013	[25]
Emomusic	Valence and arousal	song excerpts (45 s)	2013	[26]
Moodo	Mood/color-perception	music excerpts	2014	[27]
Emotify	GEMS (Geneva Emotional Music Scale) scale	song excerpts (1 min long)	2015	[28]
Amg1608	Valence and arousal	sons excerpts (30 s)	2015	[29]
GMD	Human performance and velocity annotation	audio from drum kits	2020	[30]
MERP	Valence and arousal and personal information	songs	2022	[31]

2.5 Search Results

After the full process of searching and filtering several results were obtained. During the first scope searches without the use of filters implemented by databases, the number of results was remarkably big (around 2000 results) for the given time of this project. That is why all the constraints were added and the number of papers decreased to 314 before manual filtering from all the chosen databases. The PRISMA flow chart with the exact number of papers on certain stages is presented in figure 1.

The final set of papers used in this review contains 113 documents. This is the set accomplished after applying feasibility, title and abstract filtering. The full-text reading was performed in parallel with data collection to accelerate the process of doing research. However, because of the project being individual and the short time given to finish it not all the papers from the final set were used in the information extraction process presented in 2.6. That is the final set of papers included in the analysis count 51 documents.

2.6 Data Extraction

When the final set of papers was attained we started the process of data collection (i.e. information extraction). The purpose of this step was to systematically retrieve the necessary information from the literature to answer the research questions. In this research, only one person was responsible for this work. That means one person decided on the specific data that had to be extracted from papers, read the documents and saved the found results. In addition, because of the limited time to finish this survey,

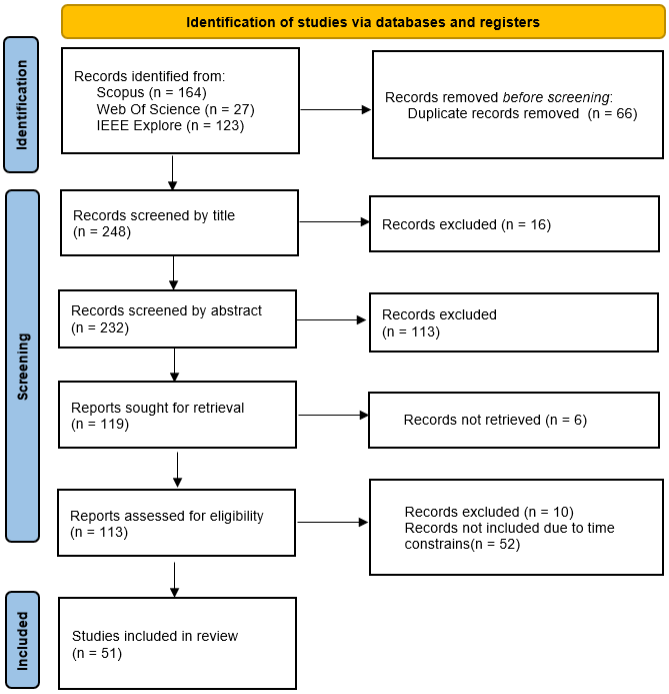

Figure 1: PRISMA flow chart

the review papers were randomly taken in batches of 10 for the data collection. Ultimately, information from 51 papers was included in the synthesis and analysis, described in 2.7.

It is critical to perform information extraction systematically for all the included papers and save the results. All the outcomes from different documents have to be comparable to each other and collectively give the data to answer research questions. That is why it was decided to define additional helper questions and save the answers to them in an Excel sheet. The specified questions along with the number of sub-question to which this question relates to are presented in table 4. In case of any missing or unclear information, it was also marked in the sheet, since missing information also gives insight into the research problem and can be taken into analysis.

Table 4: Question used in information retrieval with related sub-questions.

Question	**Related Research Sub-questions**
Publication year	5, 6
Type of input data	4
Targeted affective states	1
Which schemes were used to represent emotions and why?	2, 3, 5, 6, 7
How many schemes were used; if more than 1 why so?	2, 3, 5, 6
Is the scheme used based on psychological theory?	7
Which model type is that scheme (categorical/dimensional or both) ?	2, 3, 5, 6
Is the information retrieved from music or human reaction?	2, 5, 6
Which datasets were used?	1

2.7 Data Synthesis

The data synthesis process is one of the later parts of the systematic literature review, during which all the gathered data is analysed to formulate the results. For this specific review, the data was evaluated as one group to answer most of the sub-questions. However, for some cases, the data was grouped to enable the synthesis, such as time range groups for trend in time analysis.

After information extraction, which is in more detail described in 2.6, the needed data was collected in an Excel sheet. However, it needed some additional conversion to enable the comparison between different approaches. The data conversions and calculations that were done are:

- Conversion from specifically named schemes with a certain detail to one of five choices dimensional, categorical, both, dimensional thresholded, GEMS (Geneva Emotional Music Scale)
- Calculation of how many studies used which data input
- Calculation of how many studies used certain input data for a particular representation scheme
- Calculation how many studies used which representation scheme, where the choices were: valence/arousal dimensional scheme, GEMS (Geneva Emotional Music Scale), quadrants (derived from valence/arousal, usually happy, sad, angry, tender), emotions labelling, multi labelling, custom
- Calculation how many studies decided on a representation scheme based on psychological theories
- Grouping by years, from 2007 to 2014 included, from 2015 to 2018 included, from 2019 to 2020 included and from 2021 to 2023
- Analysis which studies used more than one representation scheme simultaneously
- Calculation how many studies had which targeted affective state and how does it influence the choice of representation scheme

The calculations and conversions presented above facilitate a precise analysis of the collected data. Furthermore, they highlight the necessary information to address the research question and sub-questions, enabling both qualitative and quantitative analysis.

3 Results

The results and answers to research sub-questions were derived from 51 papers which selection is comprehensively explained in section 2. It is important to note that the utilization of feasibility filtering may introduce bias and slightly impact the results, particularly the popularity of certain datasets may increase the popularity of an ARS used in this dataset. This section of the report is structured as follows. First in subsection 3.1 the targeted affective states are presented and explained. Subsection 3.3 describes the commonly used emotion representation schemes, analyzes their popularity of them and shows a unique simultaneous approach for the representation. The input data types are demonstrated in subsection 3.3 along with the reasoning about the connection between the input data type and the chosen representation scheme. The subsection 3.4 gives a brief analysis of the popularity of certain schemes, whereas subsection 3.5 complements the analysis with conclusions about changes in the popularity of schemes in years. Finally, subsection 3.6 describes the motivation behind choosing certain representations schemes and answers the question if researchers put attention to psychological theories while making their choices.

3.1 Targeted Affective States and Differences Between Them

This subsection will provide the answer to sub-question 1. While investigating the targeted affective states in the papers, we aim to identify what researchers are targeting and how they distinguish between different affects. An affect is a broad concept and it refers to the experience or display of emotions, moods, or feelings. Therefore, it is important to note that while emotion is a specific type of affect, it cannot be used interchangeably as a synonym for affect.

Even though the search strategy was developed in a way to include various affective states, 88% of the studies mainly focus on emotions. That is, in the majority of the papers the authors specify that their targeted affective state is the recognized or predicted emotion. Only five studies aim for mood recognition without distinct descriptions of mood as phenomena. The exact studies for certain affective states can be found in table 5. Moreover, in some studies, it is indicated if the targeted emotion is induced or perceived. The main difference between those two types is that induced emotions are intentionally

elicited and in terms of MER they are triggered by the music. Whereas perceived emotions refer to an emotion expressed by music. That is the emotions that are believed people would feel while listening to this music.

Although the difference in meaning between induced and perceived emotion is significant, a very small number of researchers clearly distinguished their target between those two. From the set of investigated papers [32] and [33] precisely define that they aim for induced emotions and [34] for perceived ones. In the rest of the studies, this distinction is not described.

Table 5: Targeted affective states and the papers that employed them.

Targeted Affective State	Records	Count
Emotion	[3, 4, 11, 32–74]	46
Mood	[75–79]	5

3.2 Types of Affect Representation Schemes

This subsection will address sub-question 2 and 3 and provide the corresponding answer. Several studies have identified various representation schemes for expressing emotions. An overview of these schemes can be found in [1], however, not all of them have been utilized in MER (Music Emotion Recognition). Generally, emotion representation schemes can be categorized into three types: *dimensional*, *categorical*, and *mixed*, which combine both dimensional and categorical approaches. Furthermore, we have identified a commonly used representation that utilizes a two-dimensional space to categorize music into four quadrants and another representation called GEMS (Geneva Emotional Music Scale). The specific studies that employed each representation scheme can be found in table 6.

Table 6: Representation schemes and the papers that employed them.

Approach	Representation scheme	Records	Count
Dimensional		[3, 4, 11, 33, 35–51, 72, 76, 78, 79]	25
Categorical	Different affects	[52–56, 68, 73–75]	8
	Quadrants	[34, 57–65, 69, 77]	12
	GEMS	[32, 66, 67]	3
Both		[70, 71]	3

Dimensional Schemes

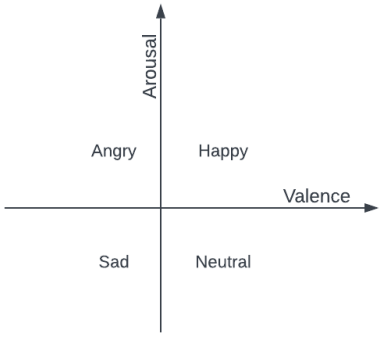

Figure 2: Dimensional emotion representation scheme.

The dimensional model, proposed initially by Russell [80] and subsequently by Thayer [81], utilizes a coordinate plane to represent emotions based on two key factors. This model portrays emotional states as points located within a two-dimensional space, characterized by valence and arousal. Valence measures the extent of positivity or negativity associated with an emotion, while arousal reflects its

intensity or level of activation. Figure 2 illustrates a visual representation of the valence/arousal emotion model.

The reasoning behind using a dimensional emotion representation scheme is diverse. Some studies claim that emotions are too complex to be represented by one label and numerical values are more reasonable and precise [38, 51]. Other researchers chose the dimensional approach because of the comparison of their work to already existing research [45, 48]. However, it is also important to note that some studies do not mention why they decided to choose this representation scheme, or they decided to follow the chosen dataset annotation approach [3, 43, 49].

Categorical Schemes

The categorical theories of emotions propose that there are universal emotions shared across cultures. In MER systems, researchers often classify emotions into four broad categories: happy, angry, sad, and neutral. These categories serve as the basis for representing and categorizing emotions within the system. In other cases, the studies propose more emotion labels like for example contentment, pleasure, disgust, surprise and others with Russell's theory as the support for this choice [74] or without a certain explanation [53]. In addition, different researchers use different names or have varying levels of granularity in defining emotions.

While investigating the existing research two specific types of categorical schemes were used apart from classes of emotions.

- *Quadrants* - This approach is based on dimensional Russell's or Thayer's scheme. However, to make the representation less complex the obtained valence/arousal values are assigned to the one quadrant made by the coordinates plane.
- *GEMS (Geneva Emotional Music Scale)* - This is a categorical model of emotion which was developed for the music domain and is based on psychological research [82]. At the top level, there are three general categories, followed by nine emotions in the middle level, and finally, 45 specific emotions at the bottom level. It is important that these emotional categories were derived from surveys that focused on the music induced emotions.

Although the schemes mentioned above are all categorical models, the reasons for using them vary. Interestingly, in many studies where different emotion classes were employed for representation, the specific reasons for their selection were often not clearly stated. Either no argumentation was provided or the researchers relied on dataset annotation as the basis for their choices. The dataset annotation was also a common reason for the use of quadrant representation, however, there were some researchers who explained their choice. For example because of popularity [34], to resolve the granularity and ambiguity of emotion labels [77] or because other studies used it [61]. On the other hand, GEMS was used since it was designed for the music domain and is based on psychological theories.

Simultaneous Use of Schemes

We have encountered 2 studies which used a unique ARS to represent affects by using dimensional and categorical representations simultaneously. The most recent study from this group is from 2023 and uses embedding space for the representation [71]. The reason behind this method is to take into consideration both the emotion categories and fine-grained discrimination and represent two data samples associated with one concept close to each other in the space.

Another research which decided to use both approaches for ARS argues that their method might provide a better capture of real-world emotions in their study [70]. They do not give a name for their scheme, however, the representation is made by weighted coefficients of a set of various components. Within this approach both the categorical and dimensional methods are applied.

3.3 Input Data

Different MER systems utilize diverse input data, which can impact the chosen emotion representation scheme. By analyzing the specific data employed, we can address sub-question 4. We identified six groups of input data: *audio samples, music feature, music features with song lyrics, music features with psychophysiological measures, EEG (Electroencephalography)* and *EEG with audio samples* and the corresponding work is presented in the table 7.

The most common input data type is the audio sample since about 65% of the studies use it. The second popular input data type is the features that include MFCCs (Mel Frequency Cepstral Coefficients), Zero Crossing Rate or timbre, harmony, frequency and more. Within this data, we can conclude that

Table 7: Input data types and papers that employed them.

Input data type	Records	Count
Audio sample	[4, 11, 32, 34, 39, 41, 42, 44, 47–49, 51–57, 59–62, 64, 67–77]	34
Music features	[3, 35–38, 40, 43, 45, 50, 63, 66, 78]	12
Music features and lyrics	[79]	1
Music features and psychophysiological measures	[46]	1
EEG (Electroencephalography)	[58, 65]	2
EEG and audio samples	[33]	1

MER research focuses on the music signal and its features to identify both the perceived and induced emotions. Moreover, it is rare to take measures from humans for MER (only 5% of analyzed studies do that) or to mix the inputs, however, still included in some research.

An interesting aspect of different input data types is the correlation of this data with the chosen emotion representation scheme. Figures 3 and 4 show the number of research which accordingly used dimensional and categorical approaches for specific inputs. From analysing the figures we realised that a majority (9 out of 12 studies) which used features as the input decided to represent the emotions in a dimensional space. From the rest of the data, there is no visible preference in the use of certain schemes for one data type since the proportions are similar or there are too few studies which use certain input data.

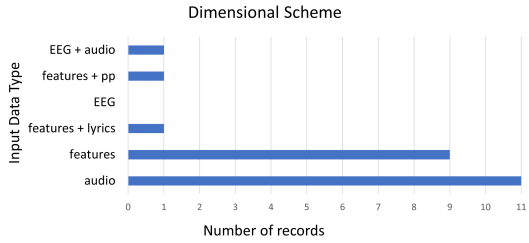

Figure 3: Number of studies which use certain types of input data and utilize dimensional ARS.

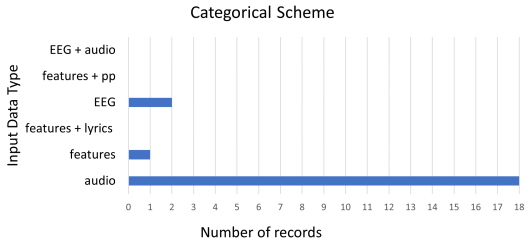

Figure 4: Number of studies which use certain types of input data and utilize categorical ARS.

3.4 Popularity of Schemes Used for Different Affective States

This subsection addresses sub-question 5 by discussing the most frequently used representation schemes for certain targeted affective states in MER. For a more detailed description of the targeted states, please refer to subsection 3.1.

Only 1 out of 51 analysed studies claimed that they work with affects in general. However, in more detail this research was focusing on emotions [3]. Other 5 papers take into consideration mood. A great majority of the studies claim that they aim to recognize or predict emotions and the specifically targeted states with according studies are shown in table 5. It is difficult to determine the popularity of certain schemes used for mood-oriented studies. The assessment is challenging because claims about one scheme being more popular than another, based on a limited set of five papers discussing mood detection, may be overstated. However, the overview of used schemes for mood-oriented research are shown in the figure 5. As presented in this figure, 3 of the studies used dimensional representation and the remaining two quadrants and categorical mood labelling.

While we can attempt to determine the popularity of the used schemes for studies targeting emotions, it is not immediately visible from the gathered data. As shown in figure 6 majority of the studies used dimensional valence/arousal representation. Moreover, it is around 44% of all the studies with the same target, which makes dimensional representation the most common. However, since it is not a majority of the used schemes we decided to not define it as a popular trend in this field.

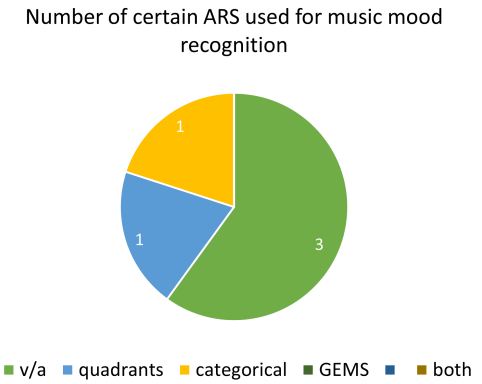

Figure 5: Number of studies which used certain affect representation schemes for mood recognition studies.

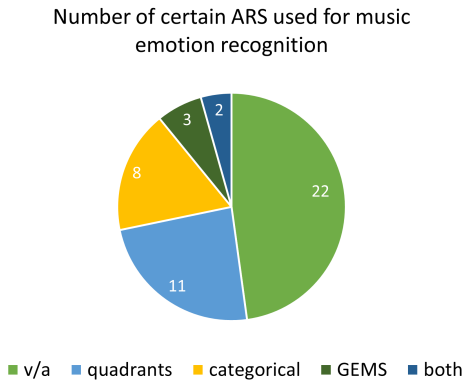

Figure 6: Number of studies which used certain affect representation schemes for emotion recognition studies.

Figure 7: v/a - valence/arousal dimensional space, quadrants - quadrant labelling based on valence/arousal space, categorical - categorical representation schemes, GEMS - Geneva Emotional Music Scales, both - use of both dimensional and categorical representation scheme

3.5 Trend Over Time in Representation Schemes Usage

This subsection aims to answer sub-question 6 by explaining how the popularity of schemes has changed over time. There exist several different types of affect representation schemes and an overview of them is described in subsection 3.3.

To derive meaningful conclusions from the analysis conducted over the years, the set of papers for this review was categorized into four different time periods. The selected time periods for analysis are as follows: 2007-2014, 2015-2018, 2019-2020, and 2021-2023. It is important to note that certain periods have different numbers of years. However, this division was determined based on the preliminary analysis of existing research and gives a nearly equal speared of the papers over the years. In general, there has been recent growth in this field, with researchers publishing an increasing number of documents each year. This is why the first time period spans seven years, while the subsequent periods cover fewer years. The last time period is not as small as the preceding, since there are not many papers published in 2023 yet.

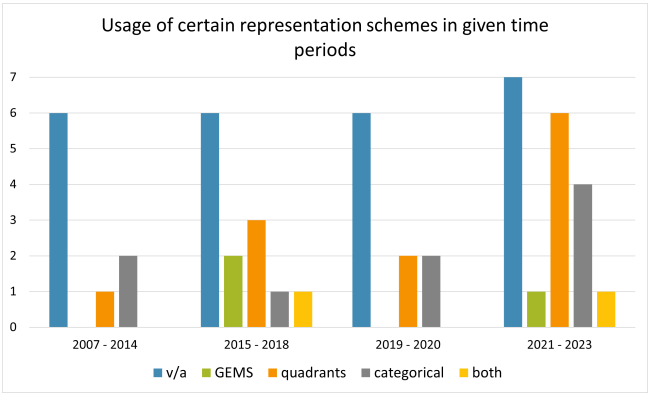

Figure 8: The number of papers which used certain representation scheme types within given time periods.

Figure 8 shows the usage of different affect representation schemes over the years. It can be realized that the usage of the dimensional valence/arousal model does not change over the years and stays constant. Which is similar to the simultaneous schemes that use both the dimensional and categorical approaches. Where it can not be said about the categorical schemes. Clearly, the utilization of this approach became more popular over the years. Since only 3 studies used GEMS representation it is not fundamental to deduce any trend in time. Another interesting finding is the unstable rise of applications of categorical approach with quadrants which is based on dimensional space.

The specifics of the utilization of certain ARS types are depicted in Figures 9 and 10. Figure 9 illustrates the count of papers that employed specific ARS types, while Figure 10 presents this count as a percentage relative to the total number of papers written in a given year. It is easily realizable that the number of written papers in this field constantly increases and even though dimensional ARS were precursors, now the amount of studies with categorical approach is close to dimensional. One interesting observation is the lack of a clear trend over the years. Researchers have been exploring different approaches, resulting in multiple crossings of lines on the figures. Furthermore, there appears to be a decrease in the year 2023. However, it is important to note that this decrease is primarily due to the limited number of papers available for that year, so it should not be considered a reliable basis for drawing conclusions.

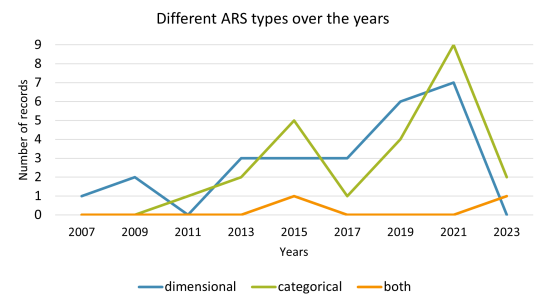

Figure 9: Number of studies which used certain types of ARS over the years.

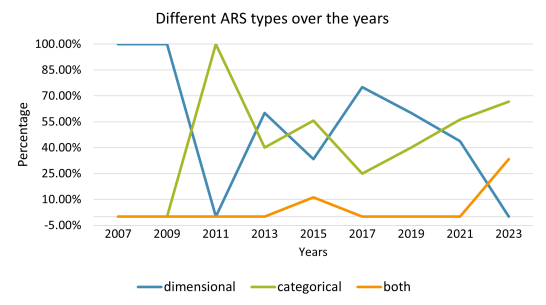

Figure 10: Percentage usage of certain type of ARS over the years.

3.6 Psychological Theories in Representation Schemes Choices

This subsection presents an answer to sub-question 7 while analysing the usage of already existing psychological research in MER. Affect is a complex concept that lacks a precise definition in psychology. However, it is crucial to base on psychological theories to represent these various affective states. The reason for this is that, while there is no consensus among experts, certain studies explore different types of affects and suggest meaningful approaches to represent them. Music affect recognition can be used in healthcare industry and aim to improve the mental health treatment. That is why making a reasonable ARS is a crucial part of those systems. When MER would be utilized in illness treatment the responsibility of proper emotional understanding will have a significant influence on humans.

While analysing the chosen papers for the review we encountered that the majority of the studies argue their choices with psychological theories. To be more precise 34 studies (approximately 66%) referred to at least one psychological work when explaining the chosen ARS and 16 did not (approximately 33%). These results might not be fully satisfactory and the MER research would be more reliable if all the studies would base on psychology.

4 Responsible Research

Performing and publishing research is a critically responsible task, which influences future studies. Innovative work can inspire future generations of researchers to follow certain paths or apply certain findings in real-world applications. That is why it is crucial to be aware of the responsibility behind the research. Section 4.1 discusses the possible ethical issues in research and argues about the possible occurrence of bias, whereas section 4.2 critically presents the importance of reproducible research.

4.1 Ethical Issues

Researchers should be conscious of the potential occurrence of various ethical problems that may arise during conducting research. While performing the review we tried to avoid creating different types of biases, as we understand their sources and potential consequences.

The first type of bias in research that we realized might occur in our process was the confirmation bias. It might take place during data interpretation. Usually, when the researchers want to validate their ideas looking for evidence in the data rather than objectively analysing it. They may do it consciously or unconsciously and in both cases it causes the subjective and wrong results. In this specific case, we wanted to find the changing popularity of certain ARS or present some interesting and groundbreaking observations. Because of that, confirmation bias might occur, which in the future may lead to wrong

decisions made by the next generation of researchers. For example, by following the trend which was not an obvious conclusion, future studies would focus more on the non-prospering methods than on the beneficial ones.

Another possible bias that may take place in performing this literature review is publishing bias. It may occur because the negative findings are less likely to be published. Then the researchers who are performing a literature review find more works with significant results and it might alter the final results and conclusions. What follows with it are inaccurate future studies which base on biased existing research.

4.2 Reproducibility of the Results

When performing research it is crucial to understand the responsibility behind it and aim to deliver reproducible work with meaningful and reliable content. Reproducibility is vital since it enhances the growth of science by letting new generation build on already existing studies. This survey was done in the possibly most systematic way and all the methods used were described to make this research replicable.

The requirements given for this research persuaded us to perform it individually with only consultations with the research group and supervisors. Because of that, all the parts of the systematic literature review were done by one person. That may lead to some bias or mistake since the chosen set of papers or extracted data was not peer-reviewed. Despite that, we applied maximum effort to make this work reproducible, by utilizing the chosen methods in a systematic way. All the decisions made during the work were motivated and in detail described to allow other researchers to replicate the research and get the same results.

Doing research in a reproducible way is critical to be convinced that the obtained results are correct and not a coincidence. That is why we documented all the choices and applied methods and this way allow other scientists to follow our path and come to the same findings.

5 Discussion and Limitations

During our research, we systematically found and analysed the ARS (Affect Representation Schemes) in studies which focused on automatic affect prediction in a music context. Affect is a complex term which is not uniquely defined and was interpreted in various ways by different researchers. The most common approach utilized by researchers was to instead of studying affects in general, target emotions and work on MER (Music Emotion Recognition).

When we compare the set of established ARS from the reviewed papers to the broad overview by [1] only a few of the possible ARS were utilized for MER. That is in music context both dimensional and categorical approaches are common and their usage is almost equally spread. However, the researchers refer to Russell's [80] and Thayer's [81] models and use only two-dimensional valence/arousal space without a third dimension. For the categorical approach we would expect the studies to mention any psychological model, for example, "Big six" postulated by Ekman [83], however, this theory was never mentioned. An interesting observation is that within this field, it is frequently observed to convert the dimensional valence/arousal model into a categorical representation. This is achieved by classifying music into one of the quadrants on a dimensional plane. This method was practised for the sake of simplicity and to resolve the granularity and ambiguity of other categorical approaches [69, 77]. Moreover, the quadrant representation is becoming more popular over the years which might cause later confusion, since it is not a method proposed by psychology, but only inspired by a psychological model. Apart from the described approaches two of the studies developed unique ARS that use the dimensional and categorical models simultaneously and both of them are based on psychological theories. Even though those methods are rare they might be the precursors for future more complex and adequate affect representation models.

In general, there is no outstanding approach in ARS for music affect prediction. Over the years researchers were trying as well the dimensional and categorical ways to represent emotions. The majority of them chose the ARS based on psychological theory, but this ratio is not satisfactory. As affects are complex and difficult to understand it is crucial to cooperate with psychologists to develop reliable and representative systems that might be later use in healthcare industry. In addition, it has to be mentioned that the feasibility filter, which was choosing the papers that mention one of the available and chosen datasets, might influence the results. Even though the used datasets are diverse in affect annotation we may have excluded relevant work for this analysis.

While performing this research we encountered some limitations that might influence the presented results. Because of the superior directives, the research had to be done within the time period of 10 weeks. That persuades us to apply additional filtering for the set of found documents to make the research feasible. What is more, all the literature searches, filtering, information extraction and analysis were performed by one person. That is also the reason why not all the papers from the final set were taken into the information extraction process. Additionally, because of the above limitations we decided not to search in ISMIR (International Society for Music Information Retrieval) Explorer. Including it in a search for this systematic review would highly excess the allocated time for searching and filtering. That is why in the final set is only the literature published by ISMIR which can be found in other included databases. In the future, we plan to include all the ISMIR publications in the review and extract the data from all papers in the final set. By doing that we would get a bigger scope and could do a more concrete analysis.

6 Conclusion and Future Work

In this work, we aimed to provide a meaningful overview of various affect representation schemes (ARS) that are used in music automatic affect prediction. Affect is a complex term which covers the expressed or observed emotional responses, however, it is often wrongly used as the equivalent of emotion, but those two differ from each other [2]. Even though we wanted to investigate research that target various affective states the majority of the reviewed studies (88 %) focused on emotion. Moreover, it is important to note that there exists a strong correlation between emotions and music [7] and because of that music emotion recognition (MER) systems have a practical application in the healthcare industry [8].

To find answers to the stated research question 1 a systematic literature review was utilized. After defining keywords and performing searches, we filtered the retrieved documents and obtained a final set of 113. Due to time limitations and the requirement that the research should be conducted by one person, 51 papers were used for data extraction. The final step involved data analysis and synthesis. Throughout the entire process, all choices made were documented and described in detail.

The ARS can be distinguished into two main categories, *dimensional* and *categorical*. The dimensional approach is based on Russel's [80] and Thayer's [81] theories and uses dimensional space to represent emotions. In this field, the researchers used only two dimensions which represented valence and arousal levels. On the other hand, the categorical approach assigns certain emotion labels in a discrete way. We have also observed different categorical ways of representations schemes, which are *GEMS (Geneva Emotional Music Scale)* developed for the music domain and *quadrants* representations which base on a dimensional approach, but assign one of the quadrants from the space to classify the music. There is no visible popularity in the ASR for music affect recognition. What is more, there is also no strict trend over the years. Even though there is more research recently, various schemes are used in a similar amount.

The researchers use different ARS and it has no meaningful correlation with the input data used for certain systems. The only conclusion that was derived is that the majority of the systems which use music features as the input data decide to represent affects in a dimensional way.

Since MER systems can be used in health care industries and improve mental health treatment [9] the psychological aspect is crucial to make the systems reliable. We have observed that only 66% of all the included studies use existing psychological research to support their choices of ARS. This implies that 33% of the studies do not refer to psychology, which, even though it represents a minority, is still significant in the context of these types of systems.

Future development is possible and would be beneficial for this research. We would suggest including the reminder papers from the final set in the data analysis part and enlarging the final set of papers with additional ones found in ISMIR (International Society for Music Information Retrieval) publications. Papers from other meaningful journals or databases might also be included. More data would increase the reliability of the results and possibly lead to some more visible trends. Another improvement could be an additional analysis of the datasets and their popularity over the years since it might influence the ARS choice. Additionally, we would recommend synthesising the data and analysing it in various ways, as there are many possibilities to find interesting results when having more time to deeply investigate the gathered data.

References

[1] M. Horvat, A. Stojanovic, and Z. Kovacevic. An overview of common emotion models in computer systems. pages 1008–1013. Institute of Electrical and Electronics Engineers Inc., 2022.

[2] Klaus R. Scherer. What are emotions? and how can they be measured?, 12 2005.

[3] Rahul Gupta, Naveen Kumar, and Shrikanth Narayanan. Affect prediction in music using boosted ensemble of filters. *2015 23rd European Signal Processing Conference, EUSIPCO 2015*, pages 11–15, 12 2015.

[4] Jia Lien Hsu, Yan Lin Zhen, Tzu Chieh Lin, and Yi Shiuan Chiu. Affective content analysis of music emotion through eeg. *Multimedia Systems*, 24:195–210, 3 2018.

[5] Lars Olov Lundqvist, Fredrik Carlsson, Per Hilmersson, and Patrik N. Juslin. Emotional responses to music: Experience, expression, and physiology. *Psychology of Music*, 37:61–90, 2009.

[6] Geneva M. Smith and Jacques Carette. What lies beneath - a survey of affective theory use in computational models of emotion. *IEEE Transactions on Affective Computing*, 13:1793–1812, 2022.

[7] Mitsuko Aramaki, Mathieu Barthet, Richard Kronland-Martinet, and Sølvi Ystad. From sounds to music and emotions, 2012.

[8] Yisi Liu, Olga Sourina, and Minh Khoa Nguyen. Real-time eeg-based emotion recognition and its applications. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6670 LNCS:256 – 277, 2011. Cited by: 184.

[9] Manasa Pisipati and Anup Nandy. Human emotion recognition using eeg signal in music listening. In *2021 IEEE 18th India Council International Conference (INDICON)*, pages 1–6, 2021.

[10] Yesid Ospitia-Medina, Sandra Baldassarri, Cecilia Sanz, José Ramón Beltrán, and José A. Olivas. Fuzzy approach for emotion recognition in music. *2020 IEEE Congreso Bienal de Argentina, ARGENCON 2020 - 2020 IEEE Biennial Congress of Argentina, ARGENCON 2020*, 12 2020.

[11] Yi Hsuan Yang, Ya Fan Su, Yu Ching Lin, and Homer H. Chen. Music emotion recognition: The role of individuality. *Proceedings of the ACM International Multimedia Conference and Exhibition*, pages 13–21, 2007.

[12] Charles Joseph and Sugeeswari Lekamge. Machine learning approaches for emotion classification of music: A systematic literature review. *2019 International Conference on Advancements in Computing, ICAC 2019*, pages 334–339, 12 2019.

[13] Deepti Chaudhary, Niraj Pratap Singh, and Sachin Singh. A survey on autonomous techniques for music classification based on human emotions recognition. *International Journal of Computing and Digital Systems*, 9:433–447, 5 2020.

[14] Matthew J Page, Joanne E McKenzie, Patrick M Bossuyt, Isabelle Boutron, Tammy C Hoffmann, Cynthia D Mulrow, Larissa Shamseer, Jennifer M Tetzlaff, Elie A Akl, Sue E Brennan, Roger Chou, Julie Glanville, Jeremy M Grimshaw, Asbjørn Hróbjartsson, Manoj M Lalu, Tianjing Li, Elizabeth W Loder, Evan Mayo-Wilson, Steve McDonald, Luke A McGuinness, Lesley A Stewart, James Thomas, Andrea C Tricco, Vivian A Welch, Penny Whiting, and David Moher. The prisma 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, page n71, 3 2021.

[15] Xu Cui, Yongrong Wu, Jipeng Wu, Zhiyu You, Jianbing Xiahou, and Menglin Ouyang. A review: Music-emotion recognition and analysis based on eeg signals. *Frontiers in Neuroinformatics*, 16:117, 10 2022.

[16] Douglas Turnbull, Luke Barrington, David Torres, and Gert Lanckriet. Semantic annotation and retrieval of music and sound effects. *IEEE Transactions on Audio, Speech and Language Processing*, 16:467–476, 2 2008.

[17] Umut Güçlü, Jordy Thielen, Michael Hanke, and Marcel A. J. van Gerven. Brains on beats. 6 2016. Tagatune.

[18] Sander Koelstra, Student Member, Christian Mü hl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, Ioannis Patras, and Senior Member. Deap: A database for emotion analysis using physiological signals.

[19] Ricardo Malheiro, Renato Panda, Paulo Gomes, and Rui Paiva. Bi-modal music emotion recognition: Novel lyrical features and dataset.

[20] R Panda, R Malheiro, B Rocha, A Oliveira, and R P Paiva. Multi-modal music emotion recognition: A new dataset, methodology and comparative analysis.

[21] Thierry Bertin-Mahieux, Daniel P W Ellis, Brian Whitman, and Paul Lamere. The million song dataset.

[22] 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.

[23] Youngmoo E Kim, Erik Schmidt, and Lloyd Emelle. Moodswings: A collaborative game for music mood label collection.

[24] Youngmoo E Kim, Erik Schmidt, and Lloyd Emelle. Moodswings: A collaborative game for music mood label collection.

[25] Jianyu Fan, Kıvanç Tatar, Miles Thorogood, and Philippe Pasquier. Ranking-based emotion recognition for experimental music.

[26] Mohammad Soleymani, Michael N. Caro, Erik M. Schmidt, Cheng Ya Sha, and Yi Hsuan Yang. 1000 songs for emotional analysis of music. pages 1–6. Association for Computing Machinery, 2013.

[27] Matevž Pesek, Primož Godec, Mojca Poredoš, Gregor Strle, Jože Guna, Emilija Stojmenova, Matevž Pogačnik, and Matija Marolt. Introducing a dataset of emotional and color responses to music.

[28] Anna Aljanaki, Frans Wiering, and Remco C. Veltkamp. Studying emotion induced by music through a crowdsourcing game. *Information Processing and Management*, 52:115–128, 1 2016.

[29] IEEE Signal Processing Society. *2015 IEEE International Conference on Acoustics, Speech and Signal Processing : proceedings : April 19-24, 2014, Brisbane Convention Exhibition Centre, Brisbane, Queensland, Australia*.

[30] Lee Callender, Curtis Hawthorne, and Jesse Engel. Improving perceptual quality of drum transcription with the expanded groove midi dataset, 2020.

[31] En Yan Koh, Kin Wai Cheuk, Kwan Yee Heung, Kat R. Agres, and Dorien Herremans. Merp: A music dataset with emotion ratings and raters' profile information. *Sensors*, 23, 1 2023.

[32] Jan Jakubik and Halina Kwasnicka Kwasnicka. Sparse coding methods for music induced emotion recognition. 2016.

[33] Kleanthis Avramidis, Christos Garoufis, Athanasia Zlatintsi, and Petros Maragos. Enhancing affective representations of music-induced eeg through multimodal supervision and latent domain adaptation. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2022-May:4588–4592, 2022.

[34] Yading Song and Simon Dixon. How well can a music emotion recognition system predict the emotional responses of participants? 2015¡br/¿.

[35] Yuchao Fan and Mingxing Xu. Mediaeval 2014: Thu-hcsil approach to emotion in music task using multi-level regression *.

[36] Konstantin Markov and Tomoko Matsui. Dynamic music emotion recognition using kernel bayes' filter. 2014.

[37] Eduardo Coutinho, Felix Weninger, Bjorn Schuller, and Klaus R. Scherer. The munich lstm-rnn approach to the mediaeval 2014 "emotion in music" task. 2014.

[38] Ye Ma, Xinxing Li, Mingxing Xu, Jia Jia, and Lianhong Cai. Multi-scale context based atention for dynamic music emotion prediction. *MM 2017 - Proceedings of the 2017 ACM Multimedia Conference*, pages 1443–1450, 10 2017.

[39] Ju-Chiang Wang, Yi-Hsuan Yang, and Hsin-Min Wang. Affective music information retrieval. pages 227–261, 2016. you can use this one for introduction ¡br/¿.

[40] Braja Gopal Patra, Promita Maitra, Dipankar Das, and Sivaji Bandyopadhyay. Mediaeval 2015: Music emotion recognition based on feed-forward neural network.

[41] Mladen Russo, Luka Kraljević, Maja Stella, and Marjan Sikora. Cochleogram-based approach for detecting perceived emotions in music. *Information Processing and Management*, 57, 9 2020. I don't understand¡br/¿¡br/¿.

[42] Pengfei Du, Xiaoyong Li, and Yali Gao. Dynamic music emotion recognition based on cnn-bilstm. 2020.

[43] Yesid Ospitia Medina, José Ramón Beltrán, and Sandra Baldassarri. Emotional classification of music using neural networks with the mediaeval dataset. *Personal and Ubiquitous Computing*, 26:1237–1249, 8 2020.

[44] Yizhuo Dong, Xinyu Yang, Xi Zhao, and Juan Li. Bidirectional convolutional recurrent sparse network (bcrsn): An efficient model for music emotion recognition. *IEEE Transactions on Multimedia*, 21:3150–3163, 12 2019.

[45] Yangyang Shu and Guandong Xu. Emotion recognition from music enhanced by domain knowledge. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11670 LNAI:121–134, 2019.

[46] Le kai Zhang, Shou qian Sun, Bai xi Xing, Rui ming Luo, and Ke jun Zhang. Using psychophysiological measures to recognize personal music emotional experience. *Frontiers of Information Technology and Electronic Engineering*, 20:964–974, 7 2019.

[47] Meixian Zhang, Yonghua Zhu, Wenjun Zhang, Yunwen Zhu, and Tianyu Feng. Modularized composite attention network for continuous music emotion recognition. *Multimedia Tools and Applications*, 2 2022.

[48] Lanqing Yin, Jiandong Tang, and Jinming Yu. Multimodal music emotion recognition based on wldnn_gan. *Proceedings - 2022 International Symposium on Advances in Informatics, Electronics and Education, ISAIEE 2022*, pages 528–532, 2022.

[49] Jinyuan Wang. Research on emotion classification of movie background music based on improved clustering algorithm. *Proceedings - 2022 11th International Conference of Information and Communication Technology, ICTech 2022*, pages 302–306, 2022.

[50] Weixin Wang. Cnn based music emotion recognition. *Proceedings - 2021 2nd International Conference on Artificial Intelligence and Computer Engineering, ICAICE 2021*, pages 190–195, 2021.

[51] Meixian Zhang, Yonghua Zhu, Ning Ge, Yunwen Zhu, Tianyu Feng, and Wenjun Zhang. Frequency embedded regularization network for continuous music emotion recognition. *Proceedings of the 2021 IEEE International Conference on Progress in Informatics and Computing, PIC 2021*, pages 426–431, 2021.

[52] Fan Zhang, Hongying Meng, and Maozhen Li. Emotion extraction and recognition from music. *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery, ICNC-FSKD 2016*, pages 1728–1733, 10 2016. they say aboit mediaeval here¡br/¿.

[53] Ja Hwung Su, Tzung Pei Hong, Yao Hong Hsieh, and Shu Min Li. Effective music emotion recognition by segment-based progressive learning. *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, 2020-October:3072–3076, 10 2020.

[54] Rajib Sarkar, Sombuddha Choudhury, Saikat Dutta, Aneek Roy, and Sanjoy Kumar Saha. Recognition of emotion in music based on deep convolutional neural network. *Multimedia Tools and Applications*, 79:765–783, 1 2020.

[55] Xiaosong Jia. A music emotion classification model based on the improved convolutional neural network. 2022.

[56] Srividya Tirunellai Rajamani, Kumar Rajamani, and Bjorn W. Schuller. Towards an efficient deep learning model for emotion and theme recognition in music. *IEEE 23rd International Workshop on Multimedia Signal Processing, MMSP 2021*, 2021. they don't say anything about emotional schemes or psychology¡br/¿.

[57] Wei Chun Chiang, Jeen Shing Wang, and Yu Liang Hsu. A music emotion recognition algorithm with hierarchical svm based classifiers. *Proceedings - 2014 International Symposium on Computer, Consumer and Control, IS3C 2014*, pages 1249–1252, 2014.

[58] Peerapon Vateekul, Nattapong Thammasan, Koichi Moriyama, Ken-Ichi Fukui, and Masayuki Numao. Item-based learning for music emotion prediction using eeg data. 2016.

[59] Zijing Gao, Beijing Key Lab, Lichen Qiu, Peng Qi, and Yan Sun. A novel music emotion recognition model for scratch-generated music. 2020.

[60] Sparsh Gupta. Deep audio embeddings and attention based music emotion recognition. *Proceedings - International Conference on Developments in eSystems Engineering, DeSE*, 2023-January:357–362, 2023.

[61] Xiao Han, Fuyang Chen, and Junrong Ban. Music emotion recognition based on a neural network with an inception-gru residual structure. *Electronics (Switzerland)*, 12, 2 2023.

[62] Jiahao Zhao, Ganghui Ru, Yi Yu, Yulun Wu, Dichucheng Li, and Wei Li. Multimodal music emotion recognition with hierarchical cross-modal attention network. *Proceedings - IEEE International Conference on Multimedia and Expo*, 2022-July, 2022.

[63] Meixian Zhang, Yonghua Zhu, Ning Ge, Yunwen Zhu, Tianyu Feng, and Wenjun Zhang. Attention-based joint feature extraction model for static music emotion classification. *Proceedings - 2021 14th International Symposium on Computational Intelligence and Design, ISCID 2021*, pages 291–296, 2021.

[64] R. Raja Subramanian, Kokkirala Aditya Ram, Dola Lokesh Sai, K. Venkatesh Reddy, Kondeti Akarsh Chowdary, and Kundu Dheeraj Datta Reddy. Deep learning aided emotion recognition from music. *International Conference on Automation, Computing and Renewable Systems, ICACRS 2022 - Proceedings*, pages 712–716, 2022. This is very weird and I think they uses their own dataset.

[65] Shengli Liao, Yumei Zhang, Honghong Yang, and Xuening Liao. Spatiotemporal emotion recognition method based on eeg signals during music listening using 1d-cnn stacked-lstm. *Proceedings - 2022 International Conference on Networking and Network Applications, NaNA 2022*, pages 7–12, 2022.

[66] Jan Jakubik and Halina Kwasnicka. Music emotion analysis using semantic embedding recurrent neural networks. *Proceedings - 2017 IEEE International Conference on INnovations in Intelligent SysTems and Applications, INISTA 2017*, pages 271–276, 8 2017.

[67] Fabio Paolizzo, Natalia Pichierri, Daniele Giardino, Marco Matta, Daniele Casali, and Giovanni Costantini. A new multilabel system for automatic music emotion recognition. *2021 IEEE International Workshop on Metrology for Industry 4.0 and IoT, MetroInd 4.0 and IoT 2021 - Proceedings*, pages 625–629, 6 2021.

[68] Aniruddha M. Ujlambkar and Vahida Z. Attar. Automatic mood classification model for indian popular music. *Proceedings - 6th Asia International Conference on Mathematical Modelling and Computer Simulation, AMS 2012*, pages 7–12, 2012.

[69] Panayu Keelawat, Nattapong Thammasan, Boonserm Kijsirikul, and Masayuki Numao. Subject-independent emotion recognition during music listening based on eeg using deep convolutional neural networks. *Proceedings - 2019 IEEE 15th International Colloquium on Signal Processing and its Applications, CSPA 2019*, pages 21–26, 4 2019.

[70] Jia Ching Wang, Yuan Shan Lee, Yu Hao Chin, Ying Ren Chen, and Wen Chi Hsieh. Hierarchical dirichlet process mixture model for music emotion recognition. *IEEE Transactions on Affective Computing*, 6:261–271, 7 2015.

[71] Naoki Takashima, Frédéric Li, Marcin Grzegorzek, and Kimiaki Shirahama. Embedding-based music emotion recognition using composite loss. *IEEE Access*, 11:36579–36604, 2023.

[72] Richard Orjesek, Roman Jarina, and Michal Chmulik. End-to-end music emotion variation detection using iteratively reconstructed deep features. *Multimedia Tools and Applications*, 81:5017–5031, 2 2022.

[73] Na He and · Sam Ferguson. Music emotion recognition based on segment-level two-stage learning. *International Journal of Multimedia Information Retrieval*, 11:383–394, 2022.

[74] Konstantin Markov and Tomoko Matsui. Music genre and emotion recognition using gaussian processes. *IEEE Access*, 2:688–697, 2014.

[75] Gaurav Agarwal and Hari Om. An efficient supervised framework for music mood recognition using autoencoder-based optimised support vector regression model. *IET Signal Processing*, 15:98–121, 4 2021. this one is long and odd and I was tired, maybe check again later.

[76] Santosh Chapaneri and Deepak Jayaswal. Structured prediction of music mood with twin gaussian processes. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10597 LNCS:647–654, 2017.

[77] Jianglong Zhang, Xianglin Huang, Lifang Yang, and Ye Xu. Some issues of mood classification for chinese popular music. *Canadian Conference on Electrical and Computer Engineering*, 2015-June:1193–1198, 6 2015.

[78] Erik M. Schmidt and Youngmoo E. Kim. Prediction of time-varying musical mood distributions using kalman filtering. *Proceedings - 9th International Conference on Machine Learning and Applications, ICMLA 2010*, pages 655–660, 2010.

[79] Chung Yi Chi, Ying Shian Wu, Wei Rong Chu, Daniel C. Wu, Jane Yung Jen Hsu, and Richard Tzong Han Tsai. The power of words: Enhancing music mood estimation with textual input of lyrics. *Proceedings - 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, ACII 2009*, 2009.

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

[81] Robert E. Thayer. The biopsychology of mood and arousal. *New York: Oxford Univ. Press*, 1989.

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

[83] Paul Ekman and Wallace V. Friesen. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 17:124–129, 1971.

Appendix

A Search Query

Search query used for literature search:

Scopus:

(TITLE-ABS-KEY (”happy” OR ”sad” OR ”angry” OR ”relaxed” OR ”affect representation schemes” OR ”affect representation” OR ”emotional states” OR ”mood” OR ”emotion*” OR ”affect*” OR ”arousal” OR ”fear” OR ”pleasure” OR ”happiness” OR ”joy” OR ”disgust” OR ”surprise” OR ”anger” OR feeling) AND TITLE-ABS-KEY (”automatic affect prediction” OR ”affect prediction” OR ”prediction” OR ”emotion recognition”) AND TITLE (”music*”) AND ALL (”RECOLA” OR ”mediaeval” OR ”GMD” OR ”DEAM” OR ”DEAP” OR ”Emotify” OR ”MoodSwings” OR ”MERP” OR ”Moodo” OR ”Amg1608” OR ”Emusic” OR ”Emomusic” OR ”MOODetector:Bi-Modal” OR ”MOODetector:Multi-Modal” OR ”cal500” OR ”msd” OR ”MagnaTagATune” OR ”gtzan” OR ”cal500exp” OR ”Soundtracks”) AND NOT TITLE (”music video” OR speech OR review OR survey)) AND (LIMIT-TO (SUBJAREA , ”COMP”)) AND (EXCLUDE (DOCTYPE , ”re”)) AND (LIMIT-TO (LANGUAGE , ”English”))

Web of Science:

”happy” OR ”sad” OR ”angry” OR ”relaxed” OR ”affect representation schemes” OR ”affect representation” OR ”emotional states” OR ”mood” OR ”emotion*” OR ”affect*” OR ”arousal” OR ”fear” OR ”pleasure” OR ”happiness” OR ”joy” OR ”disgust” OR ”surprise” OR ”anger” OR feeling (Topic) AND ”automatic affect prediction” OR ”affect prediction” OR ”prediction” OR ”emotion recognition” (Topic) AND ”music*” (Title) AND ”RECOLA” OR ”mediaeval” OR ”GMD” OR ”DEAM” OR ”DEAP” OR ”Emotify” OR ”MoodSwings” OR ”MERP” OR ”Moodo” OR ”Amg1608” OR ”Emusic” OR ”Emomusic” OR ”MOODetector:Bi-Modal” OR ”MOODetector:Multi-Modal” OR ”cal500” OR ”msd” OR ”MagnaTagATune” OR ”gtzan” OR ”cal500exp” OR ”Soundtracks” (All Fields) NOT ”music video” OR ”speech” OR ”review” OR ”survey” (Title) and Computer Science Information Systems or Computer Science Theory Methods or Computer Science Software Engineering or Computer Science Artificial Intelligence or Computer Science Cybernetics or Computer Science Interdisciplinary Applications or Computer Science Hardware Architecture (Web of Science Categories) and English (Languages)

IEEE Explore:

(”All Metadata”:”happy” OR ”All Metadata”: ”sad” OR ”All Metadata”: ”angry” OR ”All Metadata”: ”relaxed” OR ”All Metadata”: ”affect representation schemes” OR ”All Metadata”: ”affect representation” OR ”All Metadata”: ”emotional states” OR ”All Metadata”: ”mood” OR ”All Metadata”: ”emotion*” OR ”All Metadata”: ”affect*” OR ”All Metadata”: ”arousal” OR ”All Metadata”: ”fear” OR ”All Metadata”: ”pleasure” OR ”All Metadata”: ”happiness” OR ”All Metadata”: ”joy” OR ”All Metadata”: ”disgust” OR ”All Metadata”: ”surprise” OR ”All Metadata”: ”anger” OR ”All Metadata”: feeling) AND (”All Metadata”:”automatic affect prediction” OR ”All Metadata”: ”affect prediction” OR ”All Metadata”: ”prediction” OR ”All Metadata”: ”emotion recognition”) AND (”Document Title”:”music*”) AND (”Full Text Metadata”:”RECOLA” OR ”Full Text Metadata”: ”mediaeval” OR ”Full Text Metadata”: ”GMD” OR ”Full Text Metadata”: ”DEAM” OR ”Full Text Metadata”: ”DEAP” OR ”Full Text Metadata”: ”Emotify” OR ”Full Text Metadata”: ”MoodSwings” OR ”Full Text Metadata”: ”MERP” OR ”Full Text Metadata”: ”Moodo” OR ”Full Text Metadata”: ”Amg1608” OR ”Full Text Metadata”: ”Emusic” OR ”Full Text Metadata”: ”Emomusic” OR ”Full Text Metadata”: ”MOODetector:Bi-Modal” OR ”Full Text Metadata”: ”MOODetector:Multi-Modal” OR ”Full Text Metadata”: ”cal500” OR ”Full Text Metadata”: ”msd” OR ”Full Text Metadata”: ”MagnaTagATune” OR ”Full Text Metadata”: ”gtzan” OR ”Full Text Metadata”: ”cal500exp” OR ”Full Text Metadata”: ”Soundtracks”) NOT (”Document Title”:”music video” AND ”Document Title”:”speech” AND ”Document Title”:”review” AND ”Document Title”:”survey”)