Audio Engineering Society

Conference Paper

Presented at the International Conference on
Immersive and Interactive Audio
2019 March 27–29, York, UK

AI and Automatic Music Generation for Mindfulness

Duncan Williams[1], Victoria J. Hodge[1], Lina Gega[2,3], Damian Murphy[1], Peter I. Cowling[1] and Anders Drachen[1]

[1] *Digital Creativity Labs, University of York, UK.*

[2] *Dept of Health Sciences, University of York, UK.*

[3] *Hull York Medical School, University of York, UK.*

Correspondence should be addressed to Duncan Williams (duncan.williams@york.ac.uk)

ABSTRACT

This paper presents an architecture for the creation of emotionally congruent music using machine learning aided sound synthesis. Our system can generate a small corpus of music using Hidden Markov Models; we can label the pieces with emotional tags using data elicited from questionnaires. This produces a corpus of labelled music underpinned by perceptual evaluations. We then analyse participant's galvanic skin response (GSR) while listening to our generated music pieces and the emotions they describe in a questionnaire conducted after listening. These analyses reveal that there is a direct correlation between the calmness/scariness of a musical piece, the users' GSR reading and the emotions they describe feeling. From these, we will be able to estimate an emotional state using biofeedback as a control signal for a machine-learning algorithm, which generates new musical structures according to a perceptually informed musical feature similarity model. Our case study suggests various applications including in gaming, automated soundtrack generation, and mindfulness.

1 Introduction

We employ a generative music system designed to create bio-signal synchronous music in real-time according to an individual's galvanic skin response (GSR), using /machine learning (ML) techniques to determine similarity between an emotion index determined by perceptual experiment, and musical features extracted from a larger corpus of source files. This work has implications for the future design and implementation of novel portable music systems and in music-assisted mindfulness training and coaching.

1.1 Background

Mindfulness increases awareness of thoughts, feelings, and sensations, while keeping an open mind, free from distraction and judgment [1]. It can benefit mental health and general well-being [1]. The process involves the practitioner or patient concentrating their attention and awareness in a deliberate manner. Chambers [2] showed that this is correlated to galvanic skin response, heart rate variability, and the ratio of alpha and beta waves in brain imaging techniques, amongst other physiological metrics. Bondolfi [3] and Economides [1] proposed mindfulness as a therapeutic treatment and suggest there is evidence to link physiological changes with mindfulness training. Mindfulness may be considered a consolidated emotion - or more accurately, an affective state. The distinction between affective state, emotion, and mood, is complex, but in the context of sound and music, cognitive scientists have suggested that the temporal nature of the response can be a useful method of delineating between such descriptors [4].

Existing work has shown that there is a neurological and physiological connection to music [5]. When listening to our favourite music, our bodies respond physically, inducing reactions such as pupil dilation, increased heart rate, blood pressure, and skin conductivity [6]. Thus, there is a potential crossover between mindful action, physiological reaction, and musical stimulation. We are attempting to fruitfully exploit this crossover to gamify mindful interactions and create a music-based training system for the end-user using machine learning to automate the process. For example, mood-based regulation may be a target for the user. This might be adapted in the creative industries to the designs that use physiological metrics as control systems: for example in video games [7], [8] in which case, the player might be subjected to targeted mood disruption (i.e., being deliberately excited or even scared).

Machine learning (ML) is a field of computer science covering systems that learn "*when they change their behaviour in a way that makes them perform better in the future*" [9]. These systems learn from data without being specifically programmed. Many ML algorithms use supervised learning. In supervised learning, an algorithm learns a set of labelled example inputs, generates a model associating the inputs with their respective labels or scores, and then classifies (or predicts) the label or score of unseen examples using the learned model. This can emotionally label music pieces for our system.

Kim et al. [10] and Laurier & Herrera [11] give a literature overview of detecting emotion in music and focus on the music representations. Laurier & Herrera [11] also analyse the ML algorithms used. Classification algorithms used in the literature include C4.5, Gaussian mixture models, k-nearest neighbour, random forest, support vector machines, [10]–[12]. Regression techniques include Gaussian mixture model regression, multiple linear regression, partial least-squares regression and support vector regression. ML has been used to retrieve music by mood and found the personalized approach more consistent than a general approach [12]. A significant area for further work is the need to better understand the whole process and be more intelligent with respect to music, users and emotions. Thus, we underpin our system with results from human experiments. These are only feasible on a small amount of music due to the required participant sample sizes, so we augment our human-labelled data using ML. We build on these findings to deliver a personalized AI approach to target mindfulness.

1.2 Emotional responses to music

There are a number of approaches for modelling emotional responses to musical stimuli [13]. Often, these borrow from conventional models used to quantify and qualify emotion, such as the circumplex (two-dimensional) model of affect [14]. This model places valence (as a measure of positivity) and arousal (as a measure of activation strength) on the horizontal and vertical axes respectively. Emotional descriptors (e.g., happy, sad, angry) can be mapped on to this space, though some descriptors can be problematic in terms of a duality of placement on the model. For example, angry and afraid are different emotions, but would be considered negative valence and high arousal and thus difficult to differentiate on this type of emotion space.

Another problem in evaluating emotional responses to music exists in the distinction between perceived and induced emotions [15] - this is also relevant in multimodal stimulus such as film [16]. This might be broadly summarised as the listener or viewer understanding what type of feeling the stimulus is supposed to express (perceived), versus describing how it makes them feel (induced). For example, a sad piece of music may be enjoyable to an individual listener in the right context, despite being constructed to convey sadness.

2 System Overview

Recent advances in portability, wearability, and affordability of biosensors now allow us to explore evaluations considering the above distinction. Biophysiological regulation may circumnavigate some of the problems of self-reported emotion (e.g., users being unwilling to report particular felt responses, or confusing perceived responses with felt responses). Real-world testing of systems using bio-signal mappings in music generation contexts has become an emerging field of research. For example, [17] generate simple music for specific emotional states using Markov chains. The Markov chains are used to generate music while the user

wears a heart-rate sensor to monitor their bio-physiological response to the created music. The system was able to generate emotionally responsive music in a limited trial considering basic emotional descriptors.
We have developed another such system, which assumes lower skin conduction variability as a correlate for mindfulness. It attempts to generate emotionally congruent music as a training tool to promote positive affective states in the context of mindfulness. In the future, this system could also work in reverse by using skin conductance variability as a control signal to inform musical feature mapping for non-linear music generation.

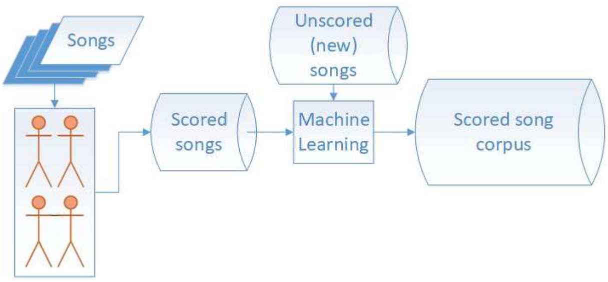

Fig. 1. Listeners rank musical excerpts, which are analysed for features to train a ML model for construction of new excerpts.

The system detects the user's current emotional level and the ML algorithm picks musical pieces to influence their future emotional level to achieve their desired mood. This whole process requires musical pieces that have an associated emotional label (score) to allow the selection of appropriate pieces. We use two tasks to achieve this. The first task in Fig. 1 and section 2.1 is to generate and expand a human-labelled corpus to provide sufficient labelled pieces for the system to operate. The music generation process is described in detail in section 2.1.1. The second task in Fig. 2 and section 2.2 is to analyse the user's galvanic skin conductivity and to select the most appropriate music from the corpus according to the user's emotional requirement. There is also a feedback loop to adapt the corpus scores according to the user's actual experience.

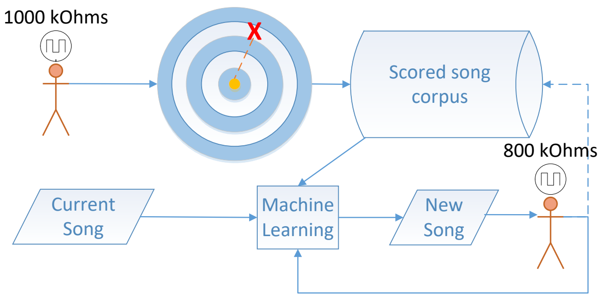

Fig. 2. We determine the change required for the user to attain their goal. The ML model selects a new piece that is musically consistent but at the required new calmness. Finally, the system has a feedback loop to remove pieces that do not influence the user's emotional level.

2.1 Task 1

To generate a database of labelled musical pieces, we elicited scores using user data accumulated through an anonymous online survey. This is an initial feasibility evaluation to assess whether human labelling is possible. Hence, we used an anonymous voluntary survey. We ran small pilot evaluations on the best survey questions and determined that binary comparison of two pieces elicited the most consistent results. We surveyed 53 participants using a Qualtrics on-line survey (www.qulatrics.com). We distributed the URL link to the survey via email lists to colleagues who responded anonymously but are English speakers, which is important for understanding the emotional labels. For this development system, we selected music that is unknown to the participants. As discussed in [11], emotions induced in the listener are influenced by many different contextual elements, such as personal experiences, cultural background, music they have only recently heard or other personal preferences, so using generated music as a stimulus may help to eliminate some of these confounds as preconceptions are removed. There is much debate regarding adjectives as emotional descriptors, and how they might be best interpreted particularly considering ambiguities across various languages. In this work we use mindful (calm/ not scary) and (tense/ scary) as these can be considered diametrically opposite on the circumplex model of affect [14]. We therefore consider "not scary vs scary" as analogous to "high vs low mindfulness" respectively, and designed the online survey asking listeners to evaluate a generated pool of training material accordingly.

Each participant evaluated four musical excerpts, 2 not scary (N1 & N2) and 2 scary excerpts (S1 & S2), in a bipolar ranking across six pairs choosing the scariest in each pair {N1vsS1, S2vsN2, S1vsN2, N1vsS2, S1vsS2 and N1vsN2}. The survey presented an initial question to allow the user to familiarize themselves with the format and then presented the six questions. The Qualtrics questionnaire allowed us to specify that each track played in full to each participant to ensure that the participant adapted fully to the track. We randomized the order of presentation of the questions (pairs of tracks) to each of the participants to reduce contextual effects. Participants were not required to answer every question in order to complete the evaluation.

2.1.1 Material

Source material was generated by training a Hidden Markov Model (HMM) and creating new permutations of the HMM with deliberate feature constraints following the procedure described in [4]. We use a transformative algorithm based on a second order Markov-model with a musical feature matrix. It allows discrete control over five parameters in a 2-dimensional model. The model is generative and can be used to create new state sequences according to the likelihood of a particular state occurring after the current and preceding states. Fig. 3 and 4 show two example scores from the stimulus set.

Fig. 3. Generated scary/angry source material.

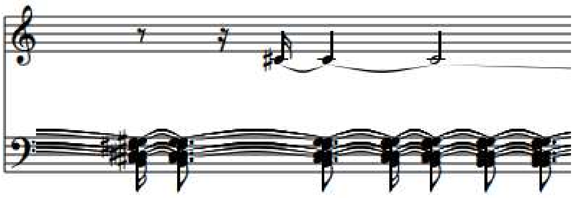

Fig. 4. Generated calm/mindful source material. Note pitch range and three # in bass clef which imply A-Maj.

2.1.2 Apparatus

Musical stimuli were rendered using a range of synthesizer timbres intended to convey the intended emotional range across an assumed emotional space.

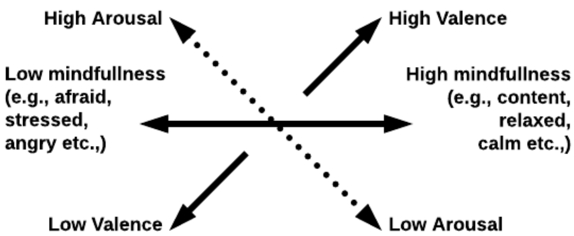

Fig..5. Chart of a rotated Valence-Arousal space after [12] bisected with a mindfulness scale assuming high mindfulness is a combination of high valence and low arousal, and vice versa for low mindfulness (with some suggested adjective labels at each end of the scale)

In this space, low mindfulness might be equated dimensionally to high arousal and low valence, or descriptively to adjectives such as scary, tense, afraid, angry, etc., whilst high mindfulness might be equated to low arousal and high valence, or adjectives like calm, content, relaxed etc. This 'mindfulness' scale can be plotted via a rotation of the traditional circumplex model of affect, as shown in Fig. 5. We generate the source MIDI files in near real-time and render them to audio with minimal latency using a DAW. Previous studies showed that the length of each music excerpt needs to be between 30 seconds to 60 seconds long to successfully induce emotions [18].All tracks were >30 seconds long, including a fade out to ensure they did not have an abrupt ending (which might otherwise also influence emotional response in the participants).

2.1.3 Results

Data from 53 participants was collected for analysis. Table 1 provides an overview of the composition of the six bipolar questions and table 2 details the participants' responses.

Table 1. Listing the pair of tracks in each question (Q1-Q6) with one question per column. Scary tracks are labelled Sn and marked with shading

Q1	Q2	Q3	Q4	Q5	Q6
N1	S2	S1	N1	S1	N1
S1	N2	N2	S2	S2	N2

Table 2. Listing the number of participants that picked each track as the scariest of the pair of tracks in each question. Each question (Q1-Q6) is one column in the table

Q1	Q2	Q3	Q4	Q5	Q6
2	37	41	5	18	24
42	5	3	41	25	20

2.1.4 Analysis

Responses to the musical stimuli in table 2 suggest that listeners found it relatively easy to discriminate the affective states between stimuli rendered using different synthesized timbres. As expected (see figures 3 and 4), shorter durations and larger pitch ranges were considered lower in mindfulness ("scarier/more tense") than longer durations with a more restricted pitch range, regardless of the timbre being used. The tracks we expected to be labelled "scary" were labelled "scary" by the participants and the tracks we expected to be labelled "not scary" were labelled "not scary". Questions 5 and 6 compare the two "scary" tracks and the two "not scary" tracks respectively. Here the results are closer as we may expect. 58.1% of participants thought S2 scarier than S1 while 54.6% felt N1 was scarier than N2.

For S1, 94.5% and 93.2% of participants rated it scarier than N1 and N2 respectively. For S2, 88.1% and 89.1% of participants rated it scarier than N1 and N2 respectively. Yet, 58.1% of the participants rated S2 scarier than S1 despite S2 having lower scariness than S1 when compared to the non-scary tracks. Similarly, for N1, 4.6% and 10.9% rated it scarier than S1 and S2 respectively while for N2, 6.8% and 11.9% rated it scarier than S1 and S2 respectively. This presents a similar contradiction as for the scary tracks as N1 has lowest scariness rating yet was rated scarier than N2 by 54.6% of participants. We cannot explain this.

Although we randomized the order of presentation of the questions to the individual participants, we did not alter the order of presentation within the questions. This may have contextual effects on the participants and needs to be considered. However, we note that the participants rated the second track as scariest in Q5 and the first track as scariest in Q6 indicating that the intra-question ordering is unlikely to be significant.

From these comparisons, we were able to attain sufficient data that we can calculate a ranked order (score) for the pieces from these pairwise comparisons [19]. From above, 58.1% of participants thought S2 scarier than S1 while 54.6% felt N1 was scarier than N2. Hence, the ranking is that S2 is scarier than S1 and N2 is calmer than N1. We produce a scored label in contrast to Laurier & Herrera [10] who used a Boolean label, for instance a song is "happy" or "not happy". However, this Boolean label does not provide the fine-grained differentiation we require to select emotionally relevant pieces so we produce a score from [0-10] for each musical piece where 0 is completely calm, 10 is completely scary and 5 is the midpoint: neither calm nor scary.

2.1.5 Enhancing the corpus

Human experiments are only feasible on a small set of music pieces as *n* pieces of music require *n!* comparisons and enough human survey participants to provide enough responses for each of the *n!* comparisons. Using human participants to generate a sufficiently large database of labelled pieces for our work is very time consuming and complex. To augment our small labelled database, we need to use ML to label new music and to provide a corpus sufficiently large for task 2 to be feasible.

2.2 Task 2

The second task is to analyse the user's galvanic skin conductance and then to select the most appropriate music. This process is similar to Huang and Cai [17] who analysed heart rate to reflect emotions and to select appropriate music pieces. Huang demonstrated varying heart rates (beats per minute (bpm)) of participants according to the emotional label of the piece played. Happy music induced the highest bpm and sad music the lowest. However, the music they labelled angry induced a very similar bpm to the music labelled joyful. We analyse skin conductivity, or galvanic skin response (GSR). When the skin's sweat glands secrete sweat, it changes the balance of positive and negative ions in the secretion, thus increasing the skin's conductivity. Measurement of GSR has been shown to be a robust metric for analysis of emotional responses to music [6], [22], [23].

The first step of task 2 requires us to compare the user's GSR signal, the emotional tag they describe after listening and the calmness level of the piece the participant is listening to. To analyze GSR, we used the Shimmer3 wireless GSR+ Unit[1] which has been validated for use in biomedical-oriented research applications1, can detect very small changes of GSR, and can stream data in real time [24]. It can also connect to recording software and export the data for extended analysis. Shimmer3 needs to be calibrated on each use through the user wearing the device for one minute to establish a baseline skin conductance signal. The baseline of each person varies due to many factors including skin dryness, nervousness (due to unfamiliarity with the experimental procedure) and ambient temperature. The captured reading for each user under analysis is their skin conductance response whilst undertaking the listening exercise, minus their individual skin conductance response baseline. After listening to each piece, the users completed a questionnaire describing the emotion they felt while listening which we compared to the GSR data [24].

[1] http://www.shimmersensing.com/products/shimmer3-wireless-GSRsensor

2.2.1 Evaluation

In Williams et al. [24], 30 participants evaluated two automatically generated pieces against two well-known pieces of scary music (the themes from the Psycho and Jaws films which were composed to be scary). The two generated pieces induced emotional responses where the response described by the participant in a questionnaire tallied with their bio-physiological responses as measured by GSR sensors. Our analyses [24] revealed that there is a direct correlation between the scariness of a musical piece, the users' GSR reading and the emotions they describe feeling in a questionnaire survey conducted after listening. Users display elevated GSR for scary pieces which they also labelled as scary in the questionnaire and lower GSR and appropriate labels for calmer pieces. Our preliminary experiments also highlighted that familiarity influences people's responses. For this reason, we focus on generating novel music to ensure that the user responds emotionally rather than responding to memories evoked by the music. We also keep to two labels: "not scary/calm" and "scary/tense" to limit confusion by reducing complexity.

This indicates that we are able to compose emotionally relevant music using HMMs. For the two film pieces, where participants were familiar with the music then the emotional responses they described in the questionnaire tallied with the expected response for a scary film but did not necessarily tally with their bio-physiological responses. For this reason, we have focussed on auto-generating music to generate our own corpus and using these to induce mindfulness and relaxation rather than selecting music from a known corpus (e.g., using streaming platforms like Spotify) which will contain pieces of music with varying degrees of familiarity for the listeners.

3 Future Work

For task 1, we will use our human-labelled data to analyse a number of ML methods to identify the best ML method with respect to accuracy foremost, but also flexibility, scalability, and adaptability.

Classification and regression algorithms need a rich data description of each piece for learning. We have developed a multi-feature music representation to enable this. We couple the symbolic musical feature data from a MIDI file, which represents the structure of the melody, chords, rhythm and other musical dimensions, with Mel-Frequency Cepstral Coefficients features [20] obtained from the entire piece to represent the piece's quality (character). This dual representation is more flexible and richer than simply using MIDI or signal-based audio content features. We only use numerical data features to describe each piece and perform feature selection to identify the most significant set of features as described in [21]. Using this reduced set of significant features, the ML model will predict the "calmness" score of new music pieces by determining the similarity between pieces using their respective sets of features.

In task 2, from the GSR analysis, we can calculate the new calmness level required to achieve a mindfulness goal (e.g., make the listener calmer if they are over-stimulated). This new calmness value allows us to retrieve all pieces of music in the scored song corpus at this new calmness level. To select the most appropriate piece from this matching set, we will match the input piece against each matching piece using the selected set of features. We will use the same music data representation as task 1 and the identical ML model to ensure consistency and to stop the system introducing contextual biasing and irregularities. We summarize task 2 in Fig. 1. The features we have selected are input to the ML model, we will calculate the musical similarity score for each piece using the ML model and then recommend the music piece that is at the correct calmness level and is most similar (musically contiguous) with the user's current state.

As we continue monitoring the participant's GSR we will assess whether the new piece has achieved the desired level of calmness. This difference (error between actual and required GSR) will feed back into the corpus of scored pieces to adjust the stimulus calmness score (essentially a calmness index). We will adjust both the global score to ensure the system correctly rates each piece and the person's own scoring mechanism to provide personalized music for their mindfulness requirements.

We can enhance the monitoring further by using additional sensors. We have proven GSR sensors for this task but other sensors such as heart-rate sensors will provide additional data. Combining data from multiple sensors will be richer. Analysing this richer data using suitable machine learning algorithms will be more accurate, more reliable and reveal finer-grained fluctuations and changes in the participant's emotional responses than would be revealed by analysing only a single sensor.

Further evaluation of our automatically generated music is not trivial. Although the influence of music on emotional state is widely acknowledged [13], [25], [26] perceptual audio evaluation strategies often consider issues of audio quality [27] rather than the influence of generative strategy on the resulting affective state in the listener. Moreover, methods which do consider the influence of generative music on affective state tend to be focused on creativity [28], and issues regarding the authorship of the material [29]. Thus, methods for perceptual evaluation of affectively-charged music generation remain a significant area for further work: for example in the design and development of a multi-purpose evaluation toolkit. Many such kits exist in audio quality evaluation, for example [30].

4 Conclusion

We have shown through an experiment with 53 participants and four music pieces that we can generate emotionally communicative music. We generated two scary pieces and two calm pieces and the users ranked these as we expected, thus supporting our hypotheses regarding how to generate calm and scary music in a more robust fashion in future.

To support this we intend to generate a larger corpus of pieces and recruit further listeners to bootstrap the generation of a larger corpus. A rich data description of human labelled pieces will allow machine-learning algorithms to label new pieces independently, which would mean we can expand the corpus to any size required for a task.

Once we have a sufficiently large labelled corpus of our auto-generated music, we will use these to select pieces to play to users according to their galvanic skin response. Our previous work [24] showed that we can combine auto-generated music and GSR monitoring to induce emotions and that these emotions correspond with those felt by the listener (as self-reported via questionnaires). The ultimate goal of the system would be to generate a calmer piece in direct response to a listener's physiological reaction and promote the necessary emotional state for enhanced mindfulness. Physiologically informed feedback is vital for this process: Any error (error between actual and required GSR) feeds back into the corpus of scored pieces to adjust that particular piece's calmness score by a small error factor. This requires a large dataset as the corpus will be adjusted gradually and incrementally to maximize the available emotion space. By using our own automatically-generated pieces, we can minimize confounds of familiarity, and the need to actively rank music whilst listening (in itself a process which might break mindfulness or relaxation). Thus the use of biophysiological sensors is critical in the development of suitable systems for audio generation in the context of mindfulness or relaxation.

Generative music technology has the potential to produce infinite soundtracks in sympathy with a listener's bio-signals, and in a biofeedback loop. There are promising applications for linking music with emotions, especially in the creative industries art and therapy, and particularly for relaxation. Enhancement of well-being using music and the emotions music induces is becoming an emerging topic for further work. The potential of cheaper wearable biosensors to collect large amounts of data for training machine learning algorithms suggests that gamifying emotions through musical sound synthesis might be possible in the near future. For example, this type of audio stimulus generation need not be restricted to a given extracted bio-signal value - in future, trials with target emotional values could be conducted, i.e., encouraging the listener to move towards a specific emotional correlate or Cartesian co-ordinate in a dimensional emotion model, such as a gamified approach to mindfulness, or a biosensor driven thriller or horror game. We note that the music generation software using HMMs allows us to generate this music rapidly so we can generate on-the-fly and on-demand in the future rather than selecting pre-generated tracks from a corpus. This auto-generation is much richer, more varied, more adaptive and more personalized than selecting from a play list.

However, such work also needs to heed the potential drawbacks of emotional manipulation using AI and related systems. There is potential for emotional manipulation for marketing purposes or social control issues. However, the promising every day applications for mindfulness, and the potential therapeutic applications of this provides a strong argument to continue investigating this area.

5 Acknowledgements

This work was supported by the Digital Creativity Labs (www.digitalcreativity.ac.uk), jointly funded by EPSRC/AHRC/Innovate UK under grant no EP/M023265/1.

References

[1] M. Economides, J. Martman, M. J. Bell, and B. Sanderson, "Improvements in Stress, Affect, and Irritability Following Brief Use of a Mindfulness-based Smartphone App: A Randomized Controlled Trial," *Mindfulness*, vol. 9, no. 5, pp. 1584–1593, Oct. 2018.

[2] R. Chambers, E. Gullone, and N. B. Allen, "Mindful emotion regulation: An integrative review," *Clin. Psychol. Rev.*, vol. 29, no. 6, pp. 560–572, 2009.

[3] G. Bondolfi, "Depression: the mindfulness method, a new approach to relapse," *Rev. Med. Suisse*, vol. 9, no. 369, p. 91, 2013.

[4] D. Williams, A. Kirke, E. R. Miranda, E. Roesch, I. Daly, and S. Nasuto, "Investigating affect in algorithmic composition systems," *Psychol. Music*, vol. 43, no.6, pp.831-854, 2014.

[5] I. Daly *et al.*, "Automated identification of neural correlates of continuous variables," *J. Neurosci. Methods*, vol. 242, pp. 65–71, 2015.

[6] S. D. Vanderark and D. Ely, "Cortisol, biochemical, and galvanic skin responses to music stimuli of different preference values by college students in biology and music," *Percept. Mot. Skills*, vol. 77, no. 1, pp. 227–234, 1993.

[7] K. Garner, "Would You Like to Hear Some Music? Music in-and-out-of-control in the Films of Quentin Tarantino," *Film Music Crit. Approaches*, pp. 188–205, 2001.

[8] M. Scirea, J. Togelius, P. Eklund, and S. Risi, "Affective evolutionary music composition with MetaCompose," *Genet. Program. Evolvable Mach.*, vol. 18, no. 4, pp. 433–465, 2017.
[9] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.
[10] Y. E. Kim *et al.*, "Music emotion recognition: A state of the art review," in *Proc. ISMIR*, 2010, pp. 255–266.
[11] C. Laurier and P. Herrera, "Automatic detection of emotion in music: Interaction with emotionally sensitive machines," in *Machine Learning: Concepts, Methodologies, Tools and Applications*, IGI Global, 2012, pp. 1330–1354.
[12] A. C. Mostafavi, Z. W. Ras, and A. Wieczorkowska, "Developing personalized classifiers for retrieving music by mood," in *Proc. Int. Workshop on New Frontiers in Mining Complex Patterns*, 2013.
[13] M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement.," *Emotion*, vol. 8, no. 4, pp. 494–521, 2008.
[14] J. A. Russell, "A circumplex model of affect.," *J. Pers. Soc. Psychol.*, vol. 39, no. 6, p. 1161, 1980.
[15] A. Gabrielsson, "Emotion perceived and emotion felt: Same or different?," *Music. Sci.*, vol. 5, no. 1 suppl, pp. 123–147, 2002.
[16] L. Tian *et al.*, "Recognizing induced emotions of movie audiences: Are induced and perceived emotions the same?," in *Affective Computing and Intelligent Interaction (ACII), 2017 Seventh International Conference on*, 2017, pp. 28–35.
[17] C.-F. Huang and Y. Cai, "Automated Music Composition Using Heart Rate Emotion Data," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, 2017, pp. 115–120.
[18] T. Eerola and J. K. Vuoskoski, "A review of music and emotion studies: approaches, emotion models, and stimuli," *Music Percept. Interdiscip. J.*, vol. 30, no. 3, pp. 307–340, 2013.
[19] F. Wauthier, M. Jordan, and N. Jojic, "Efficient ranking from pairwise comparisons," in *International Conference on Machine Learning*, 2013, pp. 109–117.
[20] B. Logan and others, "Mel Frequency Cepstral Coefficients for Music Modeling.," in *ISMIR*, 2000, vol. 270, pp. 1–11.
[21] V. J. Hodge, S. O'Keefe, and J. Austin, "Hadoop neural network for parallel and distributed feature selection," *Neural Netw.*, vol. 78, pp. 24–35, 2016.
[22] D. C. Shrift, "The galvanic skin response to two contrasting types of music," University of Kansas, Music Education, 1954.
[23] I. Daly *et al.*, "Towards human-computer music interaction: Evaluation of an affectively-driven music generator via galvanic skin response measures," in, 7th Computer Science and Electronic Engineering Conference (CEEC), IEEE, pp. 87–92, 2015.
[24] D. Williams, C-Y. Wu, V. J. Hodge, D. Murphy and P. I. Cowling, "A Psychometric Evaluation of Emotional Responses to Horror Music," in *146th Audio Engineering Society International Pro Audio Convention*, Dublin, March 20-23, 2019.
[25] K. R. Scherer, "Acoustic Concomitants of Emotional Dimensions: Judging Affect from Synthesized Tone Sequences.," in *Proceedings of the Eastern Psychological Association Meeting*, Boson, Massachusetts, 1972.
[26] K. R. Scherer, "Which Emotions Can be Induced by Music? What Are the Underlying Mechanisms? And How Can We Measure Them?," *J. New Music Res.*, vol. 33, no. 3, pp. 239–251, Sep. 2004.
[27] J. Berg and F. Rumsey, "AES E-Library: Spatial Attribute Identification and Scaling by Repertory Grid Technique and Other Methods," in *16th International Conference: Spatial Sound Reproduction*, 1999.
[28] F. Rumsey, B. de Bruyn, and N. Ford, "Graphical elicitation techniques for subjective assessment of the spatial attributes

of loudspeaker reproduction ' a pilot investigation," in *110th Audio Engineering Society Convention*, Amsterdam, 2001.

[29] O. Ben-Tal and J. Berger, "Creative aspects of sonification," *Leonardo*, vol. 37, no. 3, pp. 229–233, 2004.

[30] T. Neher, T. Brookes, and F. Rumsey, "AES Journal Forum: A Hybrid Technique for Validating Unidimensionality of Perceived Variation in a Spatial Auditory Stimulus Set," *J. Audio Eng. Soc.*, vol. 54, no. 4, pp. 259–275, 2006.