M.I.N.U.E.T.: Procedural Musical Accompaniment for Textual Narratives

Mehak Maniktala
mmanikt@ncsu.edu
North Carolina State University

Chris Miller
cwmille6@ncsu.edu
North Carolina State University

Aaron Margolese-Malin
amargol2@ncsu.edu
North Carolina State University

Arnav Jhala
ahjhala@ncsu.edu
North Carolina State University

Chris Martens
crmarten@ncsu.edu
North Carolina State University

ABSTRACT

Extensive research has been conducted on using procedural music generation in real-time applications such as accompaniment to musicians, visual narratives, and games. However, less attention has been paid to the enhancement of textual narratives through music. In this paper, we present Mood Into Note Using Extracted Text (MINUET), a novel system that can procedurally generate music for textual narrative segments using sentiment analysis. Textual analysis of the flow and sentiment derived from the text is used as input to condition accompanying music. Music generation systems have addressed variations through changes in sentiment. By using an ensemble predictor model to classify sentences as belonging to particular emotions, MINUET generates text-accompanying music with the goal of enhancing a reader's experience beyond the limits of the author's words. Music is played via the JMusic library and a set of Markov chains specific to each emotion with mood classifications evaluated via stratified 10-fold cross validation. The development of MINUET affords the reflection and analysis of features that affect the quality of generated musical accompaniment for text. It also serves as a sandbox for further evaluating sentiment-based systems on both text and music generation sides in a coherent experience of an implemented and extendable experiential artifact.

KEYWORDS

procedural content generation, music generation, mood classification, sentiment analysis, narrative experience

ACM Reference Format:
Mehak Maniktala, Chris Miller, Aaron Margolese-Malin, Arnav Jhala, and Chris Martens. 2020. M.I.N.U.E.T.: Procedural Musical Accompaniment for Textual Narratives. In *International Conference on the Foundations of Digital Games (FDG '20), September 15–18, 2020, Bugibba, Malta.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3402942.3409602

FDG '20, September 15–18, 2020, Bugibba, Malta

ACM ISBN 978-1-4503-8807-8/20/09…$15.00
https://doi.org/10.1145/3402942.3409602

1 INTRODUCTION

The belief that computational systems may one day be able to compose impactful works of music dates far back into the history of computers [13]. When it comes to artistic domains such as music and art, in particular, the idea of whether or not computers can creatively produce great works of art invariably arises and is still debated to this day[32]. In spite of this, there still exists a wide variety of ways in which computers can be designed to assist in, and guide artistic and creative processes (e.g. as a critic, as a mixed-initiative designer, etc.) [25, 48]. In this project, we explore an application of procedural music generation applied to a traditionally underrepresented domain in music generation — textual narrative. By using sentiment analysis of narrative text as input to a procedural music generator, we have developed a system for Mood Into Note Using Extracted Text (MINUET).

It should come as no surprise that text, especially narrative text, is often interwoven with emotional content. Readers have been shown to be aware of the emotional states of characters within narratives (e.g. recognizing when a character feels happy or sad) and often experience affects themselves while reading [2, 44]. This is a crucial aspect of narrative mood recognition given that authors often strive to incorporate emotional and overarching narrative elements capable of gripping the reader's attention so as to make them invested in the story's outcome [36]. Music's ability to communicate emotion and effectively designing music generation systems that produce music to do just this are both rich areas of study as well [4, 19]. With this knowledge of how emotion complements both narrative and music, we have set out to design a system in order to evaluate whether a system capable of generating music matching the mood of the text can provide additional sensory input capable of enhancing a reader's overall narrative experience. Such a system is also feasible to develop due to recent advances in sentiment analysis and procedural generation methods for a variety of content.

A great deal of our motivation stems from how naturally and intuitively emotion is incorporated into the narrative design and music composition. Yet, we are currently unaware of many other automated generative systems that strive to blend text, music, and emotion to provide new kinds of experiences to users. In this paper, we present our pilot study on developing MINUET, a system to enhance textual narrative through procedural music accompaniment. Having a system generate emotion-based music via textual analysis also affords opportunities for it to be applied to areas in which artifacts can be partially or fully described via text. One such idea involves using such a system to procedurally generate music for movies or games based on a written transcript of events. Designing such a system also presents unique challenges by the nature of the mediums involved. For example, passages in a text may involve several different moods being conveyed across multiple sentences. This presents the interesting challenge of procedurally generating music capable of blending between different moods based on the analyzed sentiments found in sequential sentences. Through our work presented here, we detail our progress on building such a system as well as offer insight into addressing interesting challenges such as these. We plan to open-source MINUET allowing the community to explore its usage in a variety of applications.

2 RELATED WORKS

2.1 Music Generation

Generally speaking, generative methods are a class of algorithms that search a space of content to output a set of acceptable solutions given input constraints and rules. Generative methods for music, for example, can search the space of all combinations of auditory pitches, timbres, and rhythms to output a subset of these combinations that sound pleasing based on rules of music theory such as keys and chords.

As Herremans et al. illustrate in [17], automated music generation has a long history within the field of computation - the variety of music generation systems and their approaches are nearly as varied as the field of music. Their taxonomy of music generation systems reflects this fact by categorizing music generation approaches whose implementations include Markov models, reinforcement learning, and even neural networks. Furthermore, these approaches can also be characterized as spanning a continuum of generative and transformational music generation algorithms as defined by Wooller in [47]. Transformational approaches in this context are generally defined by relatively small, but noticeable changes to the musical data used as the basis for generation. One such example of this approach includes altering the pitch values of notes in a piece without ultimately altering the fundamental phrase(s) of particular sections [5]. As one might expect, generative approaches instead construct music materials in a more "bottom-up" approach which expands the data size as it generates music. Namely, as opposed to transformational approaches, generative approaches do not necessarily tweak an existing musical piece and instead compose core aspects of the song from "scratch".

A somewhat similar piece of work to what we present here was accomplished by Lee and Lee [22]. Their system was developed for procedurally generating music for accompanying computer animations. Their process involved merging timings in animations with the rhythms of a piece of music by applying subtle shifts in play speeds for both. Subsequently, they utilized a directed graph of short music sequences to represent and manipulate transitions between similar musical sequences as a means of procedural music generation. Feature extraction on animations was included which yielded information such as color changes, footsteps, and arm swings. The authors mention the potential for mood features extracted from transcripts for animations, but do not include such features in any of their results. Similar to what we present here, Lee and Lee aimed to use procedurally generated music to foster deeper immersion within their presented medium. Our work presented here shifts the application domain to the text-based narrative, focuses on sentiment analysis for feature extraction, and incorporates these sentiment features as input to our procedural music generator.

Ramanto and Maulidevi [37] developed a Markov chain model that would generate a sequence of notes and then apply variations that impart an input mood upon the track. Our specific music generation approach builds upon Ramanto and Maulidevi's Markov chain to ultimately develop a *generative* procedural music generator whose generation process is closely tied to an underrepresented domain in procedural music generation—namely text-based narrative.

2.2 Emotions & Music

The strong associations between music and emotions have long been studied and considered in both the production of music designed to elicit particular emotions in listeners and in capturing a known mood in a piece of music [43]. Following this longstanding relationship between music and emotion (which is being used interchangeably with mood here), SentiMozart is a model that generates music based on mood [26]. SentiMozart uses a convolutional layer to read the emotion of a human face and then generates music designed to convey that mood by using a long short-term memory (LSTM). SentiMozart however, does not make efforts to maintain a coherent musical structure across moods. In fact, the mapping from mood to music is trained by using a different set of music pieces for each mood. Regardless, SentiMozart's unique approach towards emotion-oriented music generation inspires our work presented here.

Adam et al. [1] developed AUD.js, which provides a JavaScript API for game developers to add emotion-specific music to their web games. They used the Thayer model of mood which has two continuous dimensions, stress and energy. While their tool is intended to easily integrate into a web game and be triggered on game events, the game developers still need to specify what mood is to be evoked for which events. With MINUET, we attempt to interpret the content to determine the appropriate mood automatically.

In addition, there are many different models of mood, of which we studied only one. Our model assumes 5 basic distinguishable moods: happiness, sadness, anger, fear, and surprise. However, [37], whose Markov chain we modify, uses the Russell circumplex model of affect which has 8 different moods based on three dimensions, pleasantness, positive affect, and negative affect, as does the Tellegen-Watson-Clark model of mood, a modification which acknowledges interdependence between the dimensions of the former [24]. Meanwhile, the AUD.js music generator described earlier uses Thayer's model of mood which has only two dimensions, energy and valence. Our reason for choosing this specific model is that it is most consistent with the moods observed in the Twitter dataset.

2.3 Sentiment Analysis

Most of the research done in the field of sentiment analysis of textual data has often been based on classifying its polarity as positive, negative, and/or neutral. The authors of [46] used a prior-polarity classifier, while [33] and [14] employ naive bayes, SVM and maximum entropy. Over the past few years, RNNs and CNNs have also been used for such a binary or ternary sentiment classifier [11]. Multiclass sentiment analysis of text is a relatively new research area, with much work conducted using tweets. Mohammad and Bravo-Marquez [29] developed a regression system to identify the intensity of emotion in tweets. In [11], dos Santos and Gatti experimented with SVM and different feature-sets to identify the emotion of tweets. In this paper, we present a multi-class sentiment analysis that uses the widely available large datasets of tweets to form classifiers that can identify emotions conveyed in texts.

3 APPROACH

Our procedure for developing MINUET, a prototype system that can generate music to accompany the mood in a textual narrative involves the following two tasks: sentiment analysis of the text for mood identification, and mood-based music generation using Markov chains. We carried out sentiment analysis in Python and music generation in Java, such that they interfaced using a shared folder. Figure 1 shows a screenshot of MINUET that sequentially highlights each sentence (with different highlight colors used for different moods), displays the mood label, and plays the music accompanying the identified mood for a duration based on the number of words in a sentence. We present details on how we carried out each task as follows.

3.1 Sentiment Analysis

We used the supervised machine learning technique of multi-class classification to carry out the sentiment analysis, with five classes of mood: happy, surprise, fear, anger, and sadness. In this subsection, we present the datasets used, the preprocessing performed, the predictive models we experimented with, and our model evaluation/selection procedure.

3.1.1 Dataset. We used two labeled tweets datasets DS1 [28] and DS2 [27]. Mohammad and Bravo-Marquez in [29] describe the dataset DS1 that contains tweet entries with their content, an emotion label (joy, fear, sadness, and anger), and the emotional intensity ranging from 0 to 1. Most of the tweets with lower emotion intensity did not correlate to the emotion when inspected manually. We, therefore, removed the tweets with emotion intensities lower than 0.5. This shortened the dataset to 3534 entries with 821 labeled as joy, 834 anger, 1115 fear, and 764 sadness. We were interested in developing a model that can identify the emotion of surprise as well. So, we used another dataset DS2 that contains tweets with the following emotion tags: joy, surprise, disgust, fear, anger, sadness, with data description provided in [18, 31]. However, a high proportion of tweets were uncorrelated with their labels or misclassified. So, we manually selected 212 surprise-labeled tweets from this dataset. Our final dataset, thus, contained 3746 tweets in total with 5 classes of emotions.

3.1.2 Preprocessing. A twelve-step preprocessing was performed on the dataset: removed hyperlinks, removed words with @ (userID tags), removed hashtag symbol, expanded contractions, expanded abbreviations, removed punctuations, removed numbers, shortened elongated words (e.g. haaapppy to happy), separated joined words (caused either by a typo or hashtags), performed spell checks, removed stop words like the, is, a (to remove any bias these words can bring to the sentiment analysis), and converted the text to lowercase. Operations such as stemming or lemmatizing were not performed because they resulted in more errors than the ones they resolved. This is so because most tweets are written in informal English, and replacing words that do not exist in formal English with their stemmed or lemmatized versions was impacting the sentiment of the majority of the tweets in an unpredictable manner.

3.1.3 Predictive Models. We trained four types of classifiers on the processed dataset as explained below. We evaluated the models using both holdout and 10-fold cross validation. We performed a preliminary holdout testing by splitting the dataset into training (75%) and test sets (25%) using stratified sampling. The reason for performing holdout testing before the more robust cross validation testing is two-fold. First. we used the two best classifiers of the holdout testing to form an ensemble model as explained below. Second, we experimented with deep learning models that are known to be too computationally expensive to evaluate using cross-validation.

- *Deep Learning Models*: We developed several Deep Learning (DL) models in Python using the Keras library [15]. We experimented with different network architectures, neural network layers (LSTMs, GRUs, simple RNN layers, and convolutional layers), tuning hyperparameters (such as learning rate, optimization method, etc.), and batch sizes. We also experimented using the different types of GloVe embeddings made available by Stanford [35] in these deep learning models.
- *Naive Bayes Models*: Naive Bayes classifier [40] is a probabilistic classifier that is based on the Bayes' theorem. It follows the independence assumption between the features. We experimented with 2 variations of naive bayes classifier available in Python: GaussianNB (NB_1) in the sklearn library [34] and NaiveBayesClassifier (NB_2) in the nltk library [23].
- *Support Vector Classifier Model*: The support vector classifier (SVC) is based on support vector machines. Support vector machines [20] maximize the margin, i.e. they maximize the minimum distance between the separating hyperplane and the nearest datapoint of each class. We employed LinearSVC (one-vs-rest classifier) in the sklearn.svm library.
- *Ensemble Models*: Ensembling is a machine learning technique to aggregate the prediction of two or more models. Ensembling makes classifiers more robust and generalizable across datasets. We employed two ensemble models: a Random Forest classifier (RF) that constructs a given number of decision trees [41], and predicts the class that is the mode of the predictions of its individual trees; and Adaboost with SAMME.R ensemble algorithm (ADA) [21] on the model that performs the best in the holdout testing. We used the sklearn library to experiment with these classifiers.

3.2 Mood-based Music Generation

We used a procedural music generator to play mood-appropriate music for the text. The generator would play music of the desired

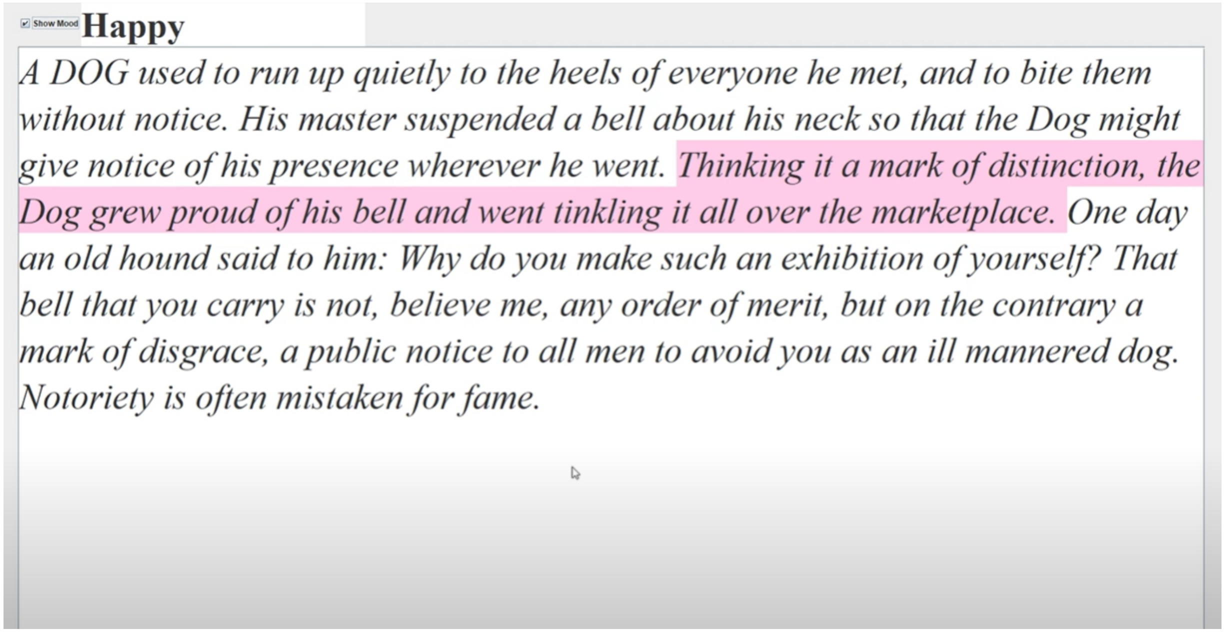

Figure 1: A screenshot of MINUET playing music for the highlighted sentence of the short story displayed. Our mood classifier identified the *Happy* mood associated with this sentence which resulted in the music generator producing music to match this mood.

mood by selecting chord and note transitions in the form of a Markov chain appropriate to that mood.

A Markov chain is a set of states each with a discrete probability distribution of transitioning to each other state including itself whenever a transition is called for. The Markov chain does not itself, however, determine when a transition is called for. It can thus be represented as an n by n square matrix where n is the number of states and each row is a probability distribution such that the probability of transitioning from state i to state j is contained in the j-th column of the i-th row. For music generation with Markov chains, it is common that each note is represented as a state, and at each beat, a new note is generated using the transition probabilities found in the Markov chain.

The music generator we used, Ramanto and Maulidevi, has one such Markov chain for each of seven common chords. For the purpose of creating mood-specific music, the generator also has a Markov chain to determine chord transitions. For instance, the chord Markov chain for sadness has higher probabilities for transitioning to minor keys that the Markov chain for happiness. Additionally, each mood has a chain for transitioning between octaves, as well as a chain that determines the duration of the next melody note (all chords are played for four beats).

It was necessary to modify Ramanto and Maulidevi's [37] music generator so that it could portray all five mood outputs of the sentiment analysis. The original music generator had distinct sets of Markov chains for surprise, sadness, and happiness, but combined anger and fear into one category they called "High Negative Affect" and described as containing distressed, fearful, hostile, jittery, nervous, and scornful emotions.

Ramanto and Maulidevi's [37] decision to combine fear and anger is not unprecedented: There are a number of papers where this is done when studying the perception of emotions in music [38, 45] because fear and anger can be evoked by very similar music, particularly when the music is limited to tempo or key or rhythm, and static in parameters like timbre and dynamics [16].

We added a new emotion to the generator with a distinct set of Markov chains from those of "High Negative Affect". We determined that fear and anger can either be passive or active and that the passive versions are distinct sounding from the active versions, but passive fear and passive anger sound similar, as do active anger and active fear. Active fear can be evoked by imagining actively running from some frightening stimulus and passive fear can be a general unease or nervousness. Active anger can be evoked as a person attacking or threatening another, while passive anger depicts a sense of internal brooding. We found that most angry tweets were threatening, suggesting external frustration. We also determined that most tweets in the fear class described nervousness - we assume that authors would be less likely to tweet in active fear, for example, while running in fear.

Using the Markov chains of fear as a place to start with, we experimented with faster tempos, octave changes, and note durations, to add a more violent and chaotic feel to the music. We found that while raising these to extremes did increase the chaos, the results would not sound like music. For instance, one experiment of making octave switches maximally rapid and chaotic was not noticeably more violent, but was less enjoyable. Instead, changing the octave chain to transition less frequently but favoring all states equally when transitioning resulted in music that was distinctly more chaotic than other emotions. The final configuration of the anger Markov chains is presented in the next section.

4 RESULTS

4.1 Sentiment Analysis

Table 2 shows the prediction accuracy for each classifier evaluated using both holdout and 10-fold cross validation (CV). Please note that we did not perform cross-validation on the DL models because it is computationally too expensive (as stated in the previous section). Additionally, the DL models did not perform well in the holdout testing, so, only the accuracy of the best DL model is presented. The configuration of this model is displayed in Table 1. In holdout testing, SVC performed better than other classifiers and so was used as the base estimator for ADA. Note that even though ADA uses SVC as the base estimator, its prediction accuracy is worse than SVC. This suggests the SVC is perhaps overfitting to the tweets dataset.

Hyperparameter	Configuration
Optimizer	Adadelta
Learning Rate	1
Decay factor	0.95
Loss Function	categorical_crossentropy
Layers	Output Shape
input_layer	(None, 21)
embedding_layer	(None, 21,50)
dropout_layer	(None, 21,50)
bidirectional_lstm_1	(None, 21,46)
bidirectional_lstm_2	(None, 128)
fully_connected_1	(None, 23)
fully_connected_2	(None, 5)

Table 1: Best Deep Learning Model's Hyperparameters

Classifier	Holdout	CV
Best DL	0.312	-
NB_1	0.637	0.586
NB_2	0.774	0.781
ADA	0.808	0.810
RF	0.847	0.841
SVC	0.877	0.861

Table 2: Prediction accuracy evaluated using holdout and 10-fold cross validation testing

Table 2 also shows the result of applying 10-fold cross validation on the classifiers. Since ADA and RF are ensemble classifiers, we observe that the decrease in prediction accuracy for these models from holdout to cross validation is of a smaller magnitude than other classifiers. We observe a similar trend of relative performance order among the classifiers in cross validation as the holdout testing, with SVC having the highest prediction accuracy. Note that the results should be read with caution because even though cross validation is a more robust method of evaluation than holdout, it cannot account of any biases that might occur within the entire dataset. Since SVC is known to overfit the data, we hypothesize that ensemble classifiers will work much better when dealing with texts from different domains. Among the ensemble models RF and ADA, RF has a better prediction accuracy than ADA, so we chose RF to classify mood for textual narrative.

4.2 Mood-based Music Generation

We performed internal testing for the new set of anger-related Markov chains. The final configuration for these set of Markov chains is shown below.

- *Tempo* - We set the tempo for the anger-based music generation to be 150 beats per minute (bpm) so that it is faster than that for low negative affects such as sad or sleepy mood, and slower than that for high engagement such as the surprise mood.
- *Chord Transition* - We modified the chord transition matrix to play music only using the minor chords, with equal probabilities of transition to each of the minor chords.
- *Note Length/Duration* - Initially, the probability of transition between any given notes was equal. We increased the probability of transition from any given note to full and half notes more than eighth and quarter notes: 16% eighth note, 20% quarter note, 32% half note, and 32% whole note
- *Octave* - Since we increased the tempo, increasing the probability of playing notes in the same octave as the previous ones improved the coherence of the music. Our final configuration for the octave transitions is as follows: 90% chance to stay in the same octave. 8% chance to shift one octave, and 2% chance to shift two octaves.

5 DISCUSSION AND FUTURE WORK

In summation, what we have presented here is a novel system for music generation based on sentiment analysis of narrative text. MINUET utilizes an ensemble model to classify sentences as belonging to particular emotions and supplies these classifications as input to a music generator that aims to generate music matching the flow of mood in text.

In the design of MINUET, we made a number of assumptions and decisions. We make the assumption that emotion can be conveyed through music [43]. For our current purposes, we also assume that the emotions are conveyed universally across listeners for a given piece of music. We acknowledge that studies have contradicted this simplification by showing that the emotions of a piece of music can be perceived differently across cultures [3] and ages [8, 43]. However, we maintain that designing an automated music generator to control for the myriad of cultures and perceptions towards music is far outside the scope of the work that we present here (although it presents several interesting questions for curious researchers). The process and methods used during the development of MINUET provide affordances for adaptation and tuning for specific preferences.

As future work, we will be conducting a series of experiments to enhance the quality of the mood classifier. In order to deal with sentences that lack in emotional intensity, we can experiment with introducing another class of "neutral" emotion. Such an approach is often employed when binary sentiment classifiers (positive and negative polarities) are not suitable for an application. We will also look into enhancing the classification capability by improving the quality of the dataset used to train the classifiers. We plan to collect more tweets (for the existing moods) using Twitter's restful API, manually tagging the tweets and evaluating the tagging. We also plan to experiment with datasets that are more geared towards narrative than tweets. Lastly, we plan to implement and experiment with the cutting edge research on attention models [39] to carry out sentiment analysis.

We also seek to improve the text narrative experience by building upon the current mood-based music generator. To make the transition of music smoother between sentences, we can experiment with the Markov chain to either include the emotion of the previous and current sentences or perform tweaks in the existing Markov chains based on the prediction probability of the identified mood. The prediction probability represents the confidence of the correctness of classification. We can use prediction probability as a measure of intensity and modify the transition table of the Markov chains based on it. Another experiment can be to determine if sentences have fragments of different sentiments, and modify both the mood classifier and music generator to adapt to such sentences as well.

With a successful prototype implemented, we now envision several interesting areas for future work and research ranging from narrative experience to emotion-based music generation. As a functional tool for authors, future studies need to include having authors use MINUET on their own (currently hindered by the pipeline having dependencies in both Java and Python libraries), and include using the pipeline on a larger more varied input text. One solution would be to implement the pipeline into a web-based application. Hosting MINUET as an open source this way would not only allow much easier access to authors looking to incorporate elements of our system into their work, but also provide a smooth, accessible interface for users looking to use MINUET as casual creators [7].

With a more streamlined and adaptive interface, improvements on MINUET can also allow it to find even more applications in mood-based music generation derived from text. Given that MINUET already allows users to input their own text, enhanced text-parsing capabilities could allow our system to also procedurally generate music for movies and games via written transcripts of events. We can also combine this prototype with generative text/story tools such as Tracery [6] or text-based interactive dramas such as Versu [12].

Similar to other studies [10, 30, 42], we would also like to allow authors to input their own melody and have the generator create variations on this melody to transform it into a specified mood. This would allow them more control over establishing consistent theme music for their story or even individual characters. We propose first using grammars or Markov chains to create single track variations of the melody, then using those variations as training data for a generative adversarial network before generating chord progressions and other tracks to accompany the variation tracks. This way the generative adversarial network can be used to approximate a human listener being able to recognize the original melody within the complex polyphonic tracks that the system generates. Research into using generative adversarial methods for music generation is currently in its infancy, but work such as MuseGAN [10] and WaveGAN [9] show promise.

REFERENCES

[1] Timothey Adam, Michael Haungs, and Foaad Khosmood. 2014. Procedurally generated, adaptive music for rapid game development. In *FDG 2014 Workshop Proceedings, Foundation of Digital Games.*

[2] Cecilia Ovesdotter Alm, Dan Roth, and Richard Sproat. 2005. Emotions from text: machine learning for text-based emotion prediction. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 579–586.

[3] Heike Argstatter. 2016. Perception of basic emotions in music: Culture-specific or multicultural? *Psychology of Music* 44, 4 (2016), 674–690.

[4] Laura-Lee Balkwill and William Forde Thompson. 1999. A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music perception: an interdisciplinary journal* 17, 1 (1999), 43–64.

[5] Karen Collins. 2009. An introduction to procedural music in video games. *Contemporary Music Review* 28, 1 (2009), 5–15.

[6] Kate Compton, Ben Kybartas, and Michael Mateas. 2015. Tracery: an author-focused generative text tool. In *International Conference on Interactive Digital Storytelling*. Springer, 154–161.

[7] Kate Compton and Michael Mateas. 2015. Casual Creators.. In *ICCC*. 228–235.

[8] Kim G Dolgin and Edward H Adelson. 1990. Age changes in the ability to interpret affect in sung and instrumentally-presented melodies. *Psychology of Music* 18, 1 (1990), 87–98.

[9] Chris Donahue, Julian McAuley, and Miller Puckette. 2018. Synthesizing Audio with Generative Adversarial Networks. *arXiv preprint arXiv:1802.04208* (2018).

[10] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang. 2018. MuseGAN: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In *Proc. AAAI Conf. Artificial Intelligence.*

[11] Cicero dos Santos and Maira Gatti. 2014. Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*. 69–78.

[12] Richard Evans and Emily Short. 2013. Versu—a simulationist storytelling system. *IEEE Transactions on Computational Intelligence and AI in Games* 6, 2 (2013), 113–130.

[13] John Fuegi and Jo Francis. 2003. Lovelace & Babbage and the creation of the 1843'notes'. *IEEE Annals of the History of Computing* 25, 4 (2003), 16–26.

[14] Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford* 1, 12 (2009).

[15] Antonio Gulli and Sujit Pal. 2017. *Deep learning with Keras*. Packt Publishing Ltd.

[16] Julia C Hailstone, Rohani Omar, Susie MD Henley, Chris Frost, Michael G Kenward, and Jason D Warren. 2009. It's not what you play, it's how you play it: Timbre affects perception of emotion in music. *The quarterly Journal of Experimental psychology* 62, 11 (2009), 2141–2155.

[17] Dorien Herremans, Ching-Hua Chuan, and Elaine Chew. 2017. A functional taxonomy of music generation systems. *ACM Computing Surveys (CSUR)* 50, 5 (2017), 1–30.

[18] David A Jurgens, Peter D Turney, Saif M Mohammad, and Keith J Holyoak. 2012. Semeval-2012 task 2: Measuring degrees of relational similarity. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation*. Association for Computational Linguistics, 356–364.

[19] Patrik N Juslin. 2001. Communicating emotion in music performance: A review and a theoretical framework. (2001).

[20] Vojislav Kecman. 2005. Support vector machines–an introduction. In *Support vector machines: theory and applications*. Springer, 1–47.

[21] Balázs Kégl. 2009. Introduction to AdaBoost.

[22] Hyun-Chul Lee and In-Kwon Lee. 2005. Automatic synchronization of background music and motion in computer animation. In *Computer Graphics Forum*, Vol. 24. Wiley Online Library, 353–361.

[23] Edward Loper and Steven Bird. 2002. NLTK: the natural language toolkit. *arXiv preprint cs/0205028* (2002).

[24] Lie Lu, Dan Liu, and Hong-Jiang Zhang. 2006. Automatic mood detection and tracking of music audio signals. *IEEE Transactions on audio, speech, and language processing* 14, 1 (2006), 5–18.

[25] Todd Lubart. 2005. How can computers be partners in the creative process: classification and commentary on the special issue. *International Journal of Human-Computer Studies* 63, 4-5 (2005), 365–369.

[26] Rishi Madhok, Shivali Goel, and Shweta Garg. 2018. SentiMozart: Music Generation based on Emotions. (2018).

[27] Saif M. Mohammad. [n.d.]. Emotion, Sentiment, and Stance Labeled Data. https://web.archive.org/web/20170623152301/http://saifmohammad.com:80/WebPages/SentimentEmotionLabeledData.html.

[28] Saif M. Mohammad. [n.d.]. Fine tuning a classifier in scikit-learn. http://saifmohammad.com/WebPages/EmotionIntensity-SharedTask.html.

[29] Saif M Mohammad and Felipe Bravo-Marquez. 2017. Emotion intensities in tweets. *arXiv preprint arXiv:1708.03696* (2017).

[30] Dan Morris, Sumit Basu, and Ian Simon. 2010. Automatic accompaniment for vocal melodies. US Patent 7,705,231.

[31] Preslav Nakov, Sara Rosenthal, Svetlana Kiritchenko, Saif M Mohammad, Zornitsa Kozareva, Alan Ritter, Veselin Stoyanov, and Xiaodan Zhu. 2016. Developing a successful SemEval task in sentiment analysis of Twitter and other social media texts. *Language Resources and Evaluation* 50, 1 (2016), 35–65.

[32] BBC News. [n.d.]. Could a computer ever create better art than a human? https://www.bbc.com/news/business-47700701.

[33] Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining.. In *LREc*, Vol. 10. 1320–1326.

[34] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in Python. *the Journal of machine Learning research* 12 (2011), 2825–2830.

[35] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543. http://www.aclweb.org/anthology/D14-1162

[36] Whitney Quesenbery and Kevin Brooks. 2010. *Storytelling for user experience: Crafting stories for better design.* Rosenfeld Media.

[37] Adhika Sigit Ramanto and Nur Ulfa Maulidevi. 2017. Markov Chain Based Procedural Music Generator with User Chosen Mood Compatibility. *International Journal of Asia Digital Art and Design Association* 21, 1 (2017), 19–24.

[38] D Ramos, JLO Bueno, and E Bigand. 2011. Manipulating Greek musical modes and tempo affects perceived musical emotion in musicians and nonmusicians. *Brazilian Journal of Medical and Biological Research* 44, 2 (2011), 165–172.

[39] Syeda Rida-E-Fatima, Ali Javed, Ameen Banjar, Aun Irtaza, Hassan Dawood, Hussain Dawood, and Abdullah Alamri. 2019. A multi-layer dual attention deep learning model with refined word embeddings for aspect-based sentiment analysis. *IEEE Access* 7 (2019), 114795–114807.

[40] Irina Rish et al. 2001. An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence*, Vol. 3. 41–46.

[41] S Rasoul Safavian and David Landgrebe. 1991. A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics* 21, 3 (1991), 660–674.

[42] Ian Simon, Dan Morris, and Sumit Basu. 2008. MySong: automatic accompaniment generation for vocal melodies. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 725–734.

[43] Swathi Swaminathan and E Glenn Schellenberg. 2015. Current emotion research in music psychology. *Emotion review* 7, 2 (2015), 189–197.

[44] Manuelde Vega. 1996. The representation of changing emotions in reading comprehension. *Cognition & Emotion* 10, 3 (1996), 303–322.

[45] Sandrine Vieillard, Isabelle Peretz, Nathalie Gosselin, Stéphanie Khalfa, Lise Gagnon, and Bernard Bouchard. 2008. Happy, sad, scary and peaceful musical excerpts for research on emotions. *Cognition & Emotion* 22, 4 (2008), 720–752.

[46] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of the conference on human language technology and empirical methods in natural language processing*. Association for Computational Linguistics, 347–354.

[47] Rene Wooller, Andrew R Brown, Eduardo Miranda, Joachim Diederich, and Rodney Berry. 2005. A framework for comparison of process in algorithmic music systems. (2005).

[48] Georgios N Yannakakis, Antonios Liapis, and Constantine Alexopoulos. 2014. Mixed-initiative co-creativity. (2014).