COVIS: Creative Orchestration of Visually-Inspired Semantics

Dylan Lasher
Computer Science
Idaho State University
Pocatello, ID. USA
lashdyla@isu.edu

Tyler Hedgepeth
Computer Science
Idaho State University
Pocatello, ID. USA
hedgtyle@isu.edu

Nickolas Nathan Taylor
Computer Science
Idaho State University
Pocatello, ID. USA
taylnick@isu.edu

***Abstract*—Extra-Inspired creative systems can learn how to usefully and creatively map and/or transform semantic concepts from remote domains to their own, at or beyond the level of human capability. This project sets out to explore how a computational system might rely on what it has learned from analyzing images to produce creative work. Specifically, the system presented here extracts semantic content and themes from an input image and uses this knowledge to creatively generate emotionally-appropriate musical artefacts to accompany the image. An independent experiment was conducted to provide a performance assessment and direct future work.**

***Index Terms*—Computational Creativity, Extra-Inspired Systems, Artificial Creativity, Music Generation, Neural Networks, CNN, RNN, K-Means, Machine Vision, Music Composition**

Github Repo: https://github.com/DLasher95/COVIS

I. Introduction

The hallmark of a creative mind is the ability to apply knowledge from outside its domain. This concept of "extra-inspired systems" allows us to take substantive inspiration from one domain and transfer it to another. Our system, for example, exists in the musical domain and extracts its inspiration from the space of visual arts. We set out to computationally bridge the gap of both these domains because, devoid of words, both music composition and imagery have the ability to posses vivid and emotional narratives.

COVIS analyzes the content of an input photographic image and uses artificial intelligence (AI) to creatively generate emotionally-equivalent music to accompany the image. The system extracts the content of the objects and actions inside the image, as well as psychology-based emotions derived from the colors present, and forms an emotional profile for the input. This way, COVIS is able to extract multiple, complicated emotional themes and map them into semantically-appropriate music audio files to accompany the image. COVIS can extract multiple, and even conflicting, emotional themes and express them in the generated artefact. We list below some of the benefits in pursuing this general line of research:

- An aid to the visually impaired, helping to appreciate the emotional magnitude conveyed by paintings, photography, and illustrations.
- A rudimentary approach to developing dynamic and appropriate soundtracks for large media, such as movies or video games.
- A tool for emotionally-rich industries, like mental health clinics, to generate calming and/or happy music without having the domain knowledge to compose music themselves.

The challenge in composing new music, just as in taking a photograph or painting a picture, is the infinite number of choices and possibilities. We present a novel approach to extract visually-semantic emotions, as well as a number of mapping rules to determine various elements of music, such as tempo or the major/minor key. Our goal with this project is to present an early framework by which the music and the visual arts can be further connected through creative AI.

II. Related Works

A. Extra-Inspired Systems

Text-to-music currently has difficulties with reliably mapping text to specific changes in sound. For this reason, prior text-to-music systems tend to rely on shallow features of text to direct generation rather than semantic context. Rangarajan, for instance, proposed three methods of mapping text to music [1].

1) Mapping letters to notes, and their frequencies to note duration.
2) Mapping vowels to notes and note duration.
3) Mapping vowels and their respective use in parts of speech to notes.

Ignoring semantic context, however, proved to be much less desirable than music generated with emotional expression in mind. Downling and Harwood [2] demonstrated the importance of tracking emotion in music. They found that copious amounts of information are processed when listening to music, and that the most impactful expressive quality that one perceives is emotion.

Davis and Mohammad took text-to-music a step further with their system *Transpose* [3]. This project set out to generate music that captured emotion dynamics in literature by studying the change in the distribution of emotion-based words. We will explore more complex emotional analytics soon, but Transpose assigned major keys to novels with more positive emotional profiles and minor keys with more negative emotional profiles. Additionally, the frequency of emotional words and the tempo of the generated music was heavily coupled. These cover a minor set of emotional parameters in music but their approach laid the groundwork for future studies on extra-inspired music generation.

Harmon [4] advanced extra-inspired systems by laying out a novel framework by which systems are able to critique the pleasantness of their own artefacts. This project performs semantic concept mining on the input text and can perform transformations on the words or phrases themselves based on its knowledge of the overall story. A search query is intelligently formed and the system stitches together ambient noises, such as rain or animal noises, in order to create atmospheric sound files. While emotional context can be gathered through this method, the sounds output could hardly be traditionally described as music.

B. Intelligent Music Composition

Automatic, or semi-automatic, music generation through computer algorithms was largely introduced to the field by Brian Eno and David Cope [5]. We also have a bedrock book on the generative theory of music authored by Lerdahl and Jackendoff [6]. This work attributed to future expansions in the automated generation of music such as Biles [7] and Collins [8]. None of these works, however, explicitly set out to capture emotion.

Downling and Harwood [2], previously mentioned, demonstrated the importance of emotions in music. Expressing emotion in music is tantamount to communicating more information for the audience to process. There has been a significant amount of work done to map emotions to discretized parameters:

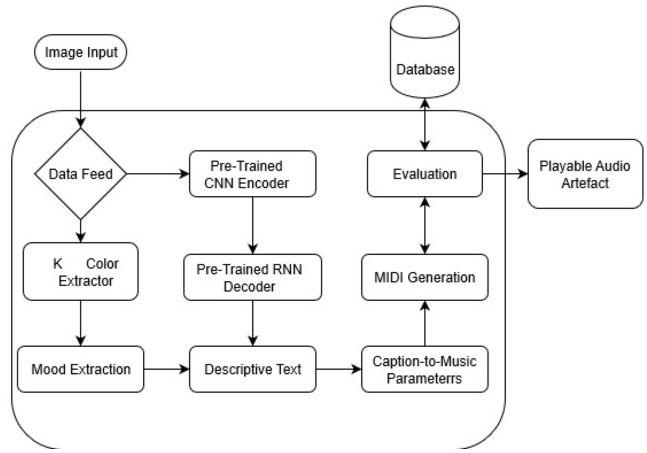
- Tempo: *Happiness* or *excitement* are attributed with fast tempos [9] [10] [11].
- Volume: *Anger*, *passion*, and *power* are attributed with high volume/intensity/loudness, whereas *fear*, *sadness*, and *weakness* are attributed with softer music [12].
- Keys: *Happiness* is attributed with major keys, whereas *sadness* is attributed with minor keys [9] [10] [11] [12].
- Melody: *Joy* and *calmness* are attributed with sequences of constant notes (constant melody), whereas *anger*, *unpleasantness*, and *excitement* are attributed with dissimilar notes (sporadic melody) [12].

It is important to note that these parameters are largely common in most cultural domains. However, an individual's unique experiences and developmental environment may influence their perception of the musical interpretation of emotions [13] [14].

III. Methods

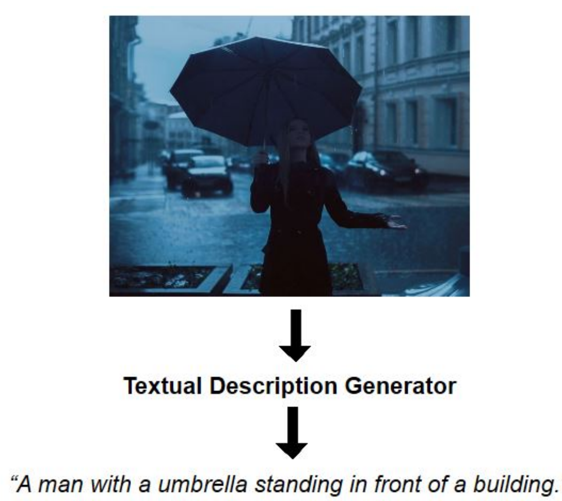
Our system, which we call COVIS, generates music according to the emotional content found in an image. It does so in several steps:

1) Analyze the input image and generate a description of the content.
2) Analyze the colors and use a color-to-emotion mapping to derive an emotional description.
3) Append the descriptions from (1) and (2) into an emotionally-descriptive profile.
4) Distill the descriptive profile by filtering out filler-words, spaces, and punctuation.
5) Feed the distilled profile into a Natural Language Processing (NLP) section that maps the words to music-writing parameters.
6) Cycle the potential artefact through an intelligent quality check until it is satisfactory.
7) Output a playable audio file artefact.



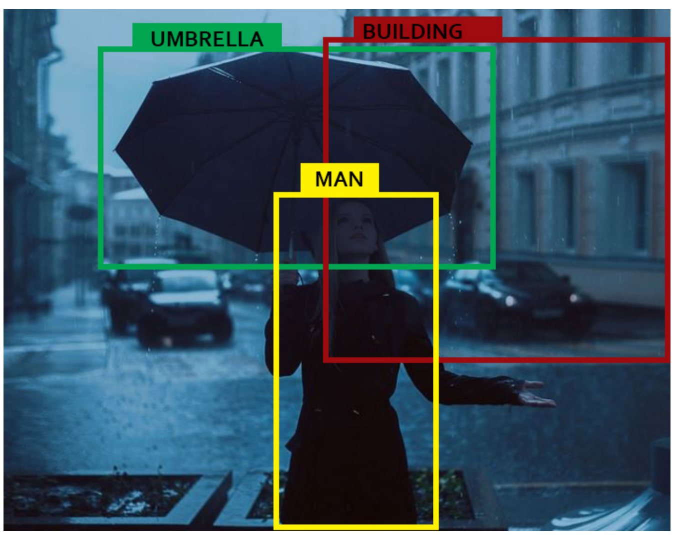
A. Automated Content Extraction

The content of an image is a strong indicator of the emotional profile of an image. Whether it be the happiness felt at a wedding or the anger depicted between two people arguing, COVIS is tasked with taking into account the objects and actions taking place in an image. Deep Learning is a rampant and expanding approach to solving problems in recent times. In this stage, COVIS extracts the content of images for later analysis. We set out to develop a Deep Learning application which combines both computer vision and natural language processing to create accurate and comprehensive captions from images provided to the system.

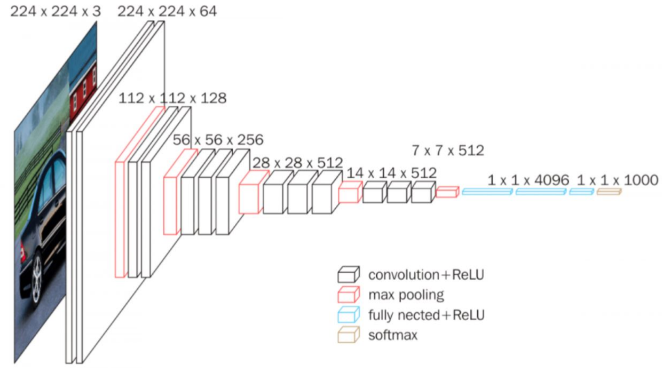


The system achieved this through two methods: A convolutional Neural Network (CNN) for extracting features out of the image and a Recurrent Neural Network (RNN) for translating the extracted features into natural sentences. We trained our system on a massive captioned set of images from image-net.org. Our approach implements the Tensorflow library and Facebook's PyTorch, which is an open-source machine learning library based on the Torch library. These are standard libraries used for applications such as computer vision and natural language processing.

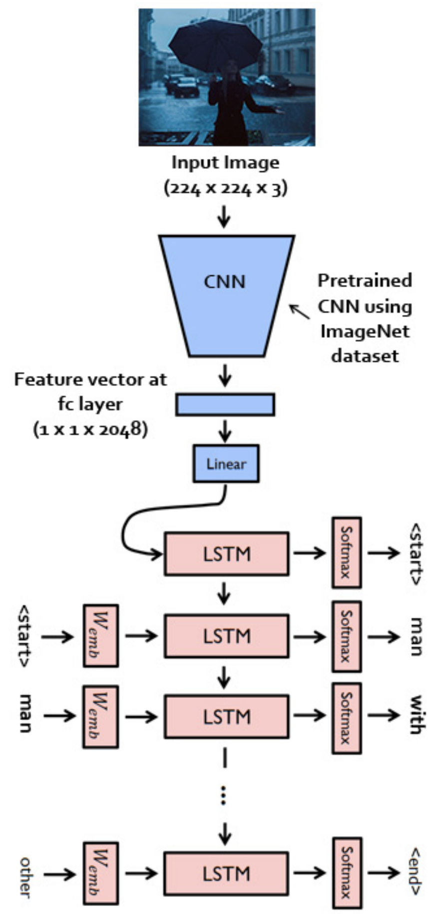
A CNN is used for feature extraction and can produce a rich representation of input images by embedding it into a fixed-length vector [17]. This representation can be used for a variety of vision tasks, which makes it a natural choice to use a CNN as an "encoder" by first pre-training it for an image classification task and using the last hidden layer as an input into an RNN. We utilized a VGG16 CNN architecture because of its overwhelming preferred use in the field of machine vision, which is outlined in Simonyan and Zisserman's [18] proposal of the system.



Because our CNN requires a specific input size (224 x 224) of an RGB pixel image, every input image must be properly resized before being fed into the CNN. These input images are then sent through the network and encoded into an array. The feature vector is linearly transformed to have the same dimension as the input dimension to the RNN network. This CNN encoding process can be seen in the following flowchart [19]:



For our RNN model, which will "decode" the CNN output vector into readable text, we chose the standard long short-term memory (LSTM) network. This is a special type of RNN which is capable of learning long-term dependencies [20]. The LSTM design is able to keep track of the objects that already have been described using text. The next words in an LSTM are based on the current time step and previous hidden state. This process continues until it gets an end-token for a sentence.



For the training phase of the RNN decoder, the pre-trained CNN extracts the feature vector from a given input image. After a linear transformation, the LSTM network input is the same as the CNN output. The decoder's source and target texts are predefined. For example, if the image description is

“A man with an umbrella standing in front of a building”, the source sequence is a list containing **[‘<start>’, ‘a’, ‘man’, ‘with’, ‘an’, ‘umbrella’, ‘standing’, ‘in’, ‘front’, ‘of’, ‘a’, ‘building’]** and the target sequence is a list containing **[‘a’, ‘man’, ‘with’, ‘an’, ‘umbrella’, ‘standing’, ‘in’, ‘front’, ‘of’, ‘a’, ‘building’, ‘<end>]**. Using these source and target sequences with the feature vector, the LSTM decoder is trained as a language model conditioned on the feature vector.

B. Color Analysis

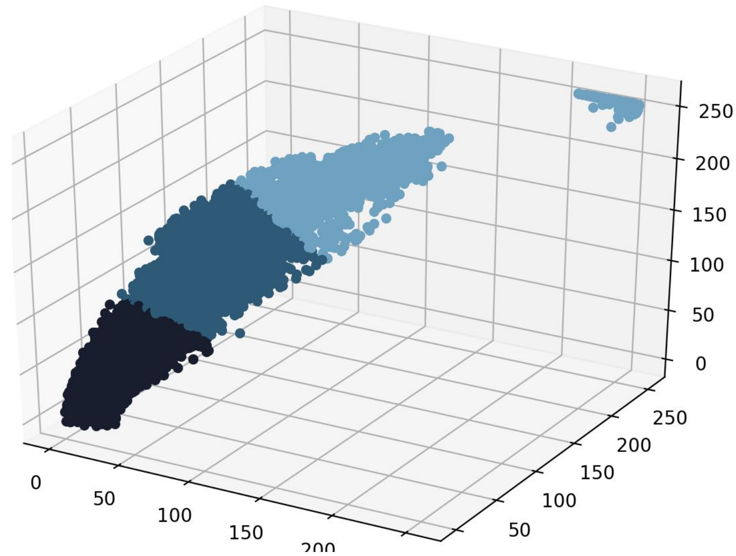
The next step of COVIS’s emotional profiling abilities lies in the color analysis of each input image. From the developmental range of childhood, emotions are commonly associated with colors. In turn, we grow up to associate colors with emotions (e.g., *anger* is “red”, *happy* is “yellow”). Extracting the most prolific colors in an image will help in establishing what colors the user is experiencing. Fugate and Franco’s [21] research into English speaker’s yielded the following guide for major RGB values:

- Anger: Red (255, 0, 0)
- Calmness: Light Blue (0,128,255), Turquoise (0, 255, 255), White (255, 255, 255)
- Disgust: Sickly Green (204, 204, 0), Orange-Brown (204, 102, 0)
- Fear: Black (0, 0, 0), Red (255, 0, 0)
- Happiness: Yellow (255, 255, 0), Turquoise (0, 255, 255)
- Sadness: Navy Blue (0, 0, 255), Gray (160, 160, 160), Gray-Blue (0, 102, 204)

To extract the dominant colors from an image, COVIS uses a k-means clustering approach. You can extract however many dominant colors “k” from the image using this method, but we chose to work with three for this study so that we may focus on a few emotions while still accounting for multiple emotional color expressions. The system uses a k-means approach because it is a powerful and unsupervised means to find dominant characteristics. It works with simple distance calculations:

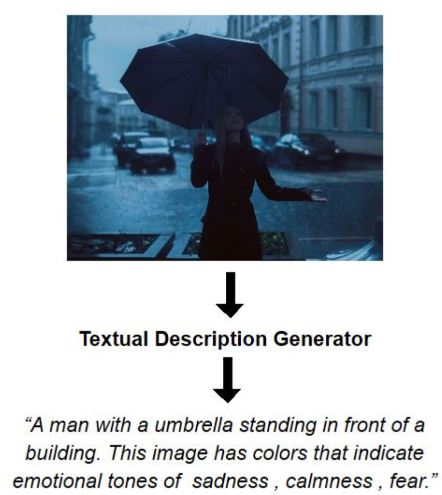
1) Initialize ‘k’ cluster centers.
2) Assign each data point to a closest cluster.
3) Compute and place the new centroid of each cluster.

We use a powerful machine learning library, scikit-learn, for our k-means analysis. The RGB color scale forms a 3-dimensional vector space with orthogonal components. We can think of each pixel as lying somewhere in the 3D color vector space. After running this algorithm on our input image, we are able to derive a portrait of the 3 clustered, dominant colors:



Once we have the RGB coordinates of our three dominant colors, we simply need to calculate the Euclidean distance to each of the Fugate-Franco [21] emotional colors. Once COVIS finds the nearest emotions to our dominant colors, we finally have our emotional color profile:

The rest of the major colors, such as violet, were not labeled with any emotion in order to avoid color values being matched to unreasonably distant emotional colors. This approach allows for the expression of multiple complex, and even disjoint, emotions that are commonly expressed through visual expressions (e.g., fear and calm). We simply have to append the emotions to the caption we previously generated:

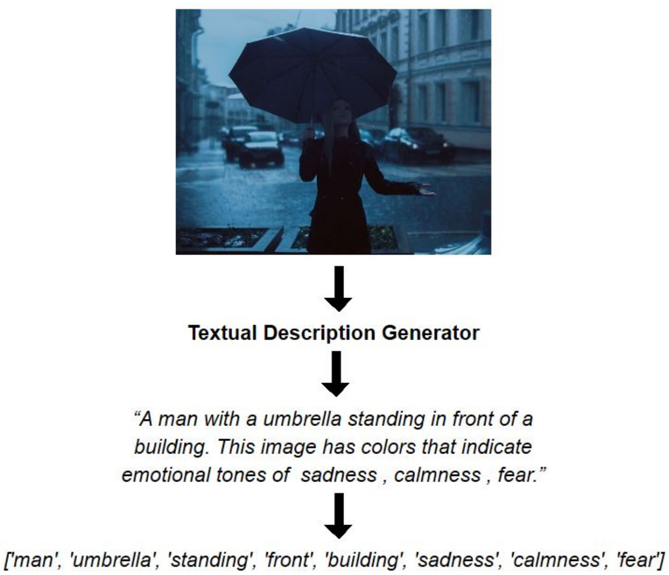


C. Caption Distillation

At this point, the system needs to hand off this semantic information to the part of the system that maps it to musical parameters. To do this, we need to convert the captioned output into a friendlier format: an array of key words. TO do this, COVIS:

1) Convert caption into an array of word strings.
2) Remove designated phrasing (e.g., *This image has colors that indicate emotional tones of "*).
3) Remove punctuation, filler words, and spaces.
4) Remove duplicate strings.

In doing so, we can better extract the key nouns and verbs that influence the overall emotional tones of the image.



IV. Music Generation and Critique

A. Semantic Evaluation

Semantic evaluation is critical to the way the system interprets and manifests the previous results. The COVIS system uses a Word2Vec approach [22] as a broad and intuitive, yet human-like way to transform input strings into meaningful musical context. An infinitely large number of musical parameters can be calculated in this way, for example: *happiness, speed, complexity*, and *pitch range*. To improve on this approach and to give the system a more full understanding of its input, antonyms can also be calculated (*slow* vs. *fast*) with a score normalized between them. A set of parameters extracted from the input text could be as follows:

Instrumentation is acquired using the same approach. By default, MIDI recognizes 128 unique instruments ranging from pianos, strings, brass, synths, and percussion. Each of these instruments is given a string representation, and an average word-distance score is calculated for each input string. The instruments with the highest scores are used in the composition.

At this point, the scores are used to interpolate between a minimum and maximum to produce real values for BPM, pitch, and mode.

Here is an example of what this section of the system produces, using input from the above section:

- Instruments: Flute, Electric Grand Piano
- Happiness score: 0.7204851752021564
- Speed score: 0.6974882928905918
- Complexity score: 0.6809697838892776
- Pitch score: 0.5133547053890055
- BPM: 138
- Mode: 4 (Mixolydian)
- Range: 49 - 83

B. Rhythm

A randomized rhythm can easily be created once a BPM (beats per minute) is established. After calculating the length of a beat, or quarter note, the following timings become trivial calculations:

- Quarter = BPM
- Eighth = Q / 2
- Triplet = Q / 3
- Sixteenth = Q / 4
- Half = Q * 2
- Whole = Q * 4

COVIS can generate a rhythm by first providing it with a length, for example, *2 * W* for a 2 measure rhythm. It then needs to know which note lengths it can use, and it will generate a sequence of note lengths which add up to the full requested length of the section, critiquing and validating each beat. This establishes a baseline for rhythmic generation which can easily be expanded upon with further development.

C. Generating Progressions

Behind every melody and harmony is a chord progression. Chord progressions provide a backbone to the melodic structure of a piece of music. COVIS defines a progression in 2 parts: *rhythm* and *intervals*. After generating a rhythm (see above), that rhythm is used to generate a sequence of intervals. While the intervals generated are essentially random, contraints are imposed on this interval generation to give a sense of intentionality, consistency, and quality. Some possible constraints include:

- The progression should prefer following a 4th, 5th, or 6th interval with the tonic to create a sense of resolution.
- The progression must either start on the tonic, or have one of the longest intervals be the tonic to reinforce the key.
- The progression must end on the tonic if it is the end of the composition to create a sense of resolution.
- Dominant chords should be followed by their relative tonic to preserve their V-I usage.
- Diminished chords should not occupy a significant amount of time within the progression.

A basic 2-5-1 jazz progression would look like this:
Rhythm: [4, 4, 8]
Intervals: [2, 5, 1]

D. Writing to MIDI

COVIS uses the Mido python library to handle the creation and editing of Midi files. COVIS takes the above generated progressions, rhythms, and melodies, and writes that information in Midi Message format that looks like this:

note_off channel=1 note=65 velocity=100 time=0.2272725

where the channel defines which instrument is playing the note, the velocity defines the volume, and the time defines the delay from the last messages. The final composition consists of a sequence of these messages.

V. Results

A. Method

To test the success of COVIS, a simple Google form was spread through ISU and several Computational Creativity related subreddits. Our survey consisted of eight examples in total with a five point Likert scale asking "how well does the music convey the emotion of the image?'.

Furthermore, this was a blind study consisting of randomly organized test samples and control samples. The four test samples had the musical artefact made by COVIS along with the image from which the artefact was generated. The other four examples, the controls, were made by pairing COVIS music pieces with the wrong picture. We included demographic metrics in the survey questions that can be used to further analyze the system's impact among different genders and levels of musical expertise.

B. Analysis

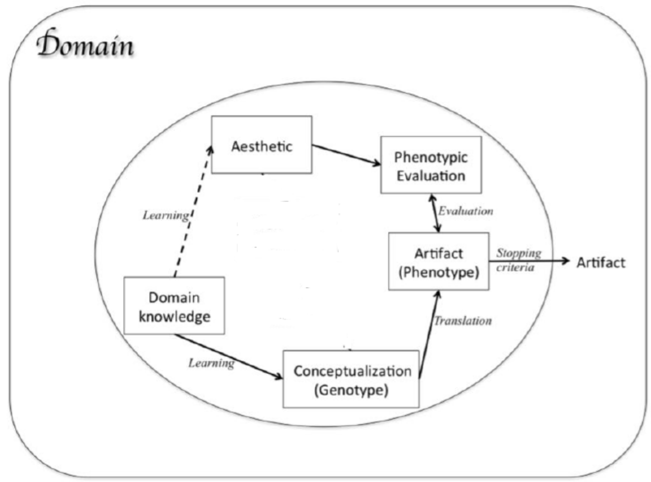
Our survey received 109 results; which are ample data to analyze. First, we sum up the times an artefact was given a four or five on each sample. We can then average the results between the two groups. We found that 60.10% of test artefacts and only 20.41% of controls were rated by participants as 'accurately portraying the emotions of the images'.

To corroborate this, a two population t-test performed on the two groups of test and control suggests a highly significant ($p < 0.0005$) difference between the test and control groups.Using the gathered gender demographic data, we tested for any difference between the success of COVIS among the different genders. We found nothing significant.

VI. Evaluation of Creativity

Beyond the ability to generate emotionally-relevant music, it is important that our system also engage in creative behavior. Prior work in computational creativity suggests that creative composition systems should strive for the highest degree of autonomy [15]. Furthermore, the system should be able to generate music that is novel and valuable for the domain audience to be considered creative [16]. We can analyze this system's creativity through two primary frameworks.

Dan Ventura [23] outlined a series of system diagrams by which to build Computationally Creative (CC) systems. This is an abstract framework incorporating a repository of domain knowledge and a domain-appropriate aesthetic that, together, inform the production of artefacts that can potentially contribute to the domain. Artefacts are represented internally as genotypic (abstract) and phenotypic (tangible) conceptualizations, eventually being translated into the final artefact.



COVIS does indeed fit a CC paradigm, where:

- **Domain**: Visual semantics and emotional music.
- **Domain Knowledge**: Humans infer contextual and emotional information from images and music must adhere to certain guidelines of semantic expression.
- **Aesthetic**: Visual psychology and musical parameters.
- **Genotype**: A semantic profile derived from an image.
- **Phenotype**: A music file with emotional context.
- **Phenotype Evaluation**: Adhering to standard frameworks of musical expression and emotional context. Also, an intelligent screening mechanism to avoid re-creating artefacts that already exist in the domain.

Here, we successfully adhere to one of Ventura's model of a proper CC system, demonstrating truly creative system behavior. However, we can analyze another framework to further corroborate the creativity of COVIS.

Pease and Colton [24] propose an approach of creative evaluation called the FACE model. They developed concrete metrics to enable us to make objective, falsifiable claims about the comparison/contrasting of different software systems. The FACE model assumes eight kinds of generative acts, in which both process (p) and artefacts (g) are produced. Every system exhibits at least one of these acts:

- $\mathbf{F}^p$: A method for generating framing information.
- $\mathbf{F}^g$: An item of framing information for $A/C/E^{p/g}$.
- $\mathbf{A}^p$: A method for generating aesthetic measures.
- $\mathbf{A}^g$: An aesthetic measure for process or product.
- $\mathbf{C}^p$: A method for generating concepts.
- $\mathbf{C}^g$: A concept.

- $\mathbf{E}^p$: A method for generating expressions of a concept.
- $\mathbf{E}^g$: An expression of a concept.

Pease and Colton [24] claim that the amount of these acts a system experiences can help determine how creative the system is being, comparatively. COVIS's FACE profile is:

- $\mathbf{F}^p$: N/A
- $\mathbf{F}^g$: Thematic context in music and images. Also, general rules of music composition.
- $\mathbf{A}^p$: Studying the domain of music composition (Manual).
- $\mathbf{A}^g$: Human-driven association of cross-thematic expression. (Manual) Also, the existence of novelty and adherence to rules of music composition (Automated).
- $\mathbf{C}^p$: Searching domain spaces until the idea formed (Manual).
- $\mathbf{C}^g$: Thematically representing emotions across visual and audio mediums (Automated).
- $\mathbf{E}^p$: Self-critiquing system for converting text into novel database-driven sound files (Automated).
- $\mathbf{E}^g$: Semantically-translated music file output (Automated).

Because we are striving for an optimally automated creative system, our "Creativity Tuple" is then $< A^g, C^g, E^p, E^g >$. Four creative acts is considered a sufficient number of acts to effectively call it "creative" [24], concluding our analysis of creativity.

In total, the creativity of COVIS can be summed up in these creative attributes:

- **Novelty**
  - COVIS is the first of its kind to extract semantic information from visual mediums and effectively translate it into musical audio.
  - Emotional expression in music is a major portion of composition, and our system is able to bridge the non-verbal gap between these domains.
- **Quality**
  - Creativity doesn't exist in a vacuum, so we must involve the audience/community [24].
  - According to our survey results, the community at large does indeed ascribe value to the artefacts produced by COVIS.
- **Typicality**
  - COVIS utilizes existing domain rules of music composition and emotional expression.
- **Intentionality**
  - COVIS sets out to extract visual semantics and uses them as an inspiration seed to generate semantically-equivalent audio files.
  - Intelligent screening process to avoid re-creating artefacts that exist in the domain.

VII. Conclusion

In this paper, we present an extra-inspired system that extracts semantic context from images and translates it into musical audio. Based on an audience survey, COVIS has empirically demonstrated that it is able to usefully and creatively map and/or transform semantic concepts between two separate domains.

For future steps towards improving the system, we would like to experiment further with thematic extraction techniques (i.e. extract more elements from an image for semantic analyses). Additionally, we believe that integrating models that are able to learn and express more advanced musical semantic expression techniques by studying existing music, such as an evolutionary algorithm. Training the thematic extractor on non-photographic images would also expand its descriptive abilities, broadening the range of inspiration sets. Lastly, we believe that COVIS can be thought of as a preliminary support tool for music generation. Here, we implemented general models of musical expression, but users have the full range to specify further musical parameters. Users would be able to integrate rules of jazz, for instance, in order to exclusively express semantic jazz music.

It is important to also specify that COVIS has successfully been evaluated as computationally creative software under two separate frameworks of CC. The artefacts COVIS produces are not mere generations of musical audio, but artificially thoughtful and creative expressions of emotional imagery. It is the authors' hope that the next steps toward COVIS-inspired systems will help to further enrich inspired music generation.

References

[1] Rangarajan, R. 2015. Generating music from natural language text. In *Digital Information Management (ICDIM), 2015 Tenth International Conference on, 85–88*. IEEE.

[2] W. Jay Dowling and Dane L Harwood. 1986. *Music cognition*, volume 19986. Academic Press New York.

[3] Davis, H., and Mohammad, S. M. 2014. Generating music from literature. *arXiv preprint arXiv:1403.2124*.

[4] Harmon, S. (2017). Narrative-inspired Generation of Ambient Music. ICCC.

[5] David Cope. 1996. *Experiments in musical intelligence*, volume 12. AR Editions Madison, WI.

[6] Fred Lerdahl and Ray S Jackendoff. 1983. *A generative theory of tonal music.* MIT press.

[7] John Biles. 1994. Genjam: A genetic algorithm for generating jazz solos. pages 131–137.

[8] Nick Collins. 2008. The analysis of generative music programs. *Organised Sound*, 13(3):237–248.

[9] Patrick G Hunter, E Glenn Schellenberg, and Ulrich Schimmack. 2010. Feelings and perceptions of happiness and sadness induced by music: Similarities, differences, and mixed emotions. *Psychology of Aesthetics, Creativity, and the Arts*, 4(1):47.

[10] S Omar Ali and Zehra F Peynirciolu. 2010. Intensity of emotions conveyed and elicited by familiar and unfamiliar music. *Music Perception: An Interdisciplinary Journal*, 27(3):177–182.

[11] Gregory D Webster and Catherine G Weir. 2005. Emotional responses to music: Interactive effects of mode, texture, and tempo. *Motivation and Emotion*, 29(1):19–39.

[12] Alf Gabrielsson and Erik Lindstr¨om. 2001. The influence of musical structure on emotional expression.

[13] 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*, pages 43–64.

[14] Steven J Morrison and Steven M Demorest. 2009. Cultural constraints on music perception and cognition. *Progress in brain research*, 178:67–77.

[15] Wiggins, G. A.; Pearce, M. T.; M¨ullensiefen, D.; et al. 2009. *Computational modeling of music cognition and musical creativity.*

[16] Boden, M. A. 2009. Computer models of creativity. *AI Magazine* 30(3):23.
[17] Oriol Vinayals and Alexander Toshev and Samy Bengio and Dumitru Erhan. (2014). Show and Tell: A Neural Image Caption Generator. *arXiv:1411.4555*.
[18] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv preprint arXiv:1409.1556*.
[19] Xiangyu Zhang and Jianhua Zou and Kaiming He and Jian Sun. 2015. Accelerating Very Deep Convolutional Networks for Classification and Detection. *arXiv:1505.06798*.
[20] Hochreiter, Sepp Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation. 9. 1735-80. 10.1162/neco.1997.9.8.1735.
[21] Fugate, J., Franco, C. L. (2019). What Color Is Your Anger? Assessing Color-Emotion Pairings in English Speakers. *Frontiers in psychology*, 10, 206. https://doi.org/10.3389/fpsyg.2019.00206
[22] US9740680B1 - Computing numeric representations of words in a high-dimensional space. (n.d.). Retrieved from https://patents.google.com/patent/US9740680B1/en
[23] Ventura, D. (2017). How to Build a CC System. *ICCC*.
[24] Colton, Simon Charnley, John Pease, Alison. (2012). Computational creativity theory: The FACE and IDEA descriptive models. Proceedings of the 2nd International Conference on Computational Creativity, ICCC 2011.