Generating audiovisual summaries from literary works using emotion analysis

Daniela F. Milon-Flores*, Jose Ochoa-Luna* and Erick Gomez-Nieto†
*Department of Computer Science, Universidad Católica San Pablo, Arequipa, Peru
† Center for the Study of Violence, University of São Paulo, São Paulo-SP, Brazil

***Abstract*—Literature work reading is an essential activity for human communication and learning. However, several relevant tasks as selection, filter or analyze in a high number of such works become complex. For dealing with this requirement, several strategies are proposed to rapidly inspect substantial amounts of text, or retrieve information previously read, exploiting graphical, textual or auditory resources. In this paper, we propose a methodology to generate audiovisual summaries by the combination of emotion-based music composition and graph-based animation. We applied natural language processing algorithms for extracting emotions and characters involved in literary work. Then, we use the extracted information to compose a musical piece to accompany the visual narration of the story aiming to convey the extracted emotion. For that, we set important musical features as chord progressions, tempo, scale, and octaves, and we assign a set of suitable instruments. Moreover, we animate a graph to sum up the dialogues between the characters in the literary work. Finally, to assess the quality of our methodology, we conducted two user studies that reveal that our proposal provides a high level of understanding over the content of the literary work besides bringing a pleasant experience to the user.**

I. Introduction

Nowadays, the Digital Age that we live provides to us a wide set of multimedia resources in support to accomplish routine tasks. All of these resources aim to enhance human capabilities that we naturally are impaired to perform, *e.g.*, video calls, massive information search, share photos instantly, ask online for delivery food, *etc*. In particular, some of these have been addressed to transfer meaning or emotions from text in books to music [1] and visual metaphors [2]. On one hand, affective algorithm composition (AAC) [3] is a field that combines computer aided composition and emotional assessment. This involves works connecting automatic detection of emotions with the automatic generation of music. On the other hand, there were several attempts to visually understand and analyze text in books [4]. Topology networks graphs, charts, timelines, to name a few, have been used to better characterize and summarize text in novels and poems [5]–[7]. However, to the best of our knowledge there are no works that automatically compose emotion-based music and provide visual insights from text at the same time. The aim of this work is to fill this gap, and to offer a better experience when reading books, mostly novels.

In this sense, our contribution is twofold. First, we propose a methodology to generate *audiovisual summaries* by the combination of emotion-based music composition and graph-based animation. Second, we provide a prototype implementation which allow us to evaluate the audiovisual summaries created.

In order to generate audiovisual summaries, we first applied natural language processing algorithms for extracting emotions and characters involved in literary work. Then, guided by those emotions, we compose a musical piece to accompany the visual narration of the story. It is worth noting that music and visual summary are delivered at the same time. In the affective algorithmic composition, important musical features like chord progressions, tempo, scale, and octaves have been considered. In addition, a set of suitable instruments for conveying the emotions extracted have been selected.

To create a visual summary, we develop a graph-based animation for representing how characters interact among them. In fact, we count and display variation frequencies of each character along with the story by using graph nodes. The relationships among characters are represented by edges — the width of the edge denotes the frequency of interaction between two specific characters.

The prototype implementation allow us to evaluate our setup. In particular, we were interested in assessing its summarization and perceptive conveying capabilities. Our experiments on several literature novels reveal that our proposal provides a high level of understanding on the content of a specific book's section. Besides, it brings a pleasant experience to the user.

The remainder of the paper is organized as follows. Related works are presented in Section 2. In Section 3 our methodology to produce audiovisual summaries is presented. Section 4 is devoted to results. Finally, Section 5 concludes the paper.

II. Related Work

To the best of our knowledge there is no previous works combining automated musical composition and summary visualization from text at the same time. Thus, in this section, we review separately related works to automatic musical composition and visual summarization of texts.

A. Automated Musical Composition

As stated by Williams et al. [3], affective algorithm composition (AAC) is a field that combines computer-aided composition and emotional assessment. These authors presented an overview of this field, including a classification of such systems by feature-set, seed material and control data. Our

proposal can be considered an AAC system where emotion detection and composition are performed automatically.

In this context, to tackle this task we first need to detect emotions in text and then generate affective algorithms for music composition. In order to perform emotion detection, several natural language processing algorithms, mainly based on sentiment analysis [8], can be used.

Several works in text-to-speech synthesis have used emotion detection to produce speech consistent with the emotions in the text [9]. However, there are few works connecting automatic detection of emotions with the automatic generation of music. We review two main works that address this issue, and therefore are the more related to ours.

Transpose [1] is a software that sonifies a novel using the connection between emotions and various musical elements, such as scale or tempo. Based on the text, they generate a piano composition for conveying several feelings encountered in the book. Using a lexicon they mapped words from a text into one of eight different emotions, such as *fear*, *joy*, *anger*, *surprise*. Then, they generate the octave, scale, and tempo by using extracted emotions and create note sequences for each section. Finally, these notes and values are put into JFugue[1] a Java API for music generation. This application combines all this features resulting in an audio file.

In the work of Stere and Trausan-Matu [10] a poem is analyzed for generating a piano composition whose sound is in harmony with the mood transmitted by the poem. Three aspects are considered: rhythm (words stressed and unstressed), punctuation (punctuation marks), and mood (*joyful* or *sadness*). The rhythm of the song is divided into four-quarter notes measures. Then, each sequence of note durations is mapped to an appropriate musical note. In order to detect the mood of the poem they use a Naive Bayes classifier. They also use JFugue API for creating the musical composition.

While not directly related to the two previous works, the system Robin [11] offers an interactive tool for generating music based on a certain type of representation of emotions. In fact, the emotions are used as an interactive metaphor that allows the users to control the music. In order to do so, the user should communicate his/her emotions to Robin, the algorithmic composer that interprets this information and reconfigures the composition.

Recently deep neural networks demonstrated their ability in learning from big data collections that generate music by neural networks [12]. In our setting a deep neural network approach would not be useful, because the composition is guided by emotions. Another issue would be to construct a training dataset larger enough to avoid over-fitting in such a complex models.

B. Visual Summarization of Texts

In this section we review works that maps emotions/meaning in text to visual metaphors. A proposal of visualization of literary work is presented by Yamada and Murai [5]. By detecting and counting keywords in Shakespeare's plays they capture the meaning of a story, i.e., its structural pattern. Furthermore, they create a visual distribution of the keywords to extract the outline of the history. To do so, the structural pattern is expressed by a continuous scalar distribution. Then, the numerical distribution is converted to a color picture using an original color chart that is relevant to the story. In this sense, the text is visualized by visual imagery. This initial work was further explored in [4]. Story patterns in novels were visualized using density analysis. Keywords in multiple aspects were selected to deal with multiple story-lines in parallel. A complex narrative can be simultaneously visualized with multiple themes.

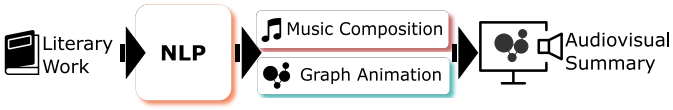


Fig. 1. The pipeline used for our proposed methodology.

The work presented by Mohammad [6] uses a set of techniques that automatically access and analyze books (mostly novels and fairy tales) using sentiment analysis in tandem with effective visualizations. This allows one to quantify and track emotions in both individual books and across very large collections. Visualization of emotions is performed using pie charts, timelines and word clouds.

In our approach, we perform visual summarization based on topological analysis. Therefore the two following works are the most related to ours. Elson et al [2] show a method aimed to extract social networks from literature. The network is constructed based on dialogue interaction between characters in a novel. In this setting, characters are vertices and edges signify an amount of bilateral conversation between those characters. Edge weights corresponds to the frequency and length of their exchanges. To do so, the authors propose a novel combination of pattern-based detection, statistical methods, and adaptation of standard natural language tools for the literacy genre.

Another topological analysis is performed by Waumans et al. [7]. They construct social networks based on the processing of dialogs in the text of novels. The characters intervening in each conversations are identified, and a network is formed between them based on these interactions. The end result is a signature of each story, characterizing elements such as the scope, the number of protagonists and even the author's reading level.

III. Our approach

A. Music Theory Preliminaries

Following, we briefly describe some fundamentals of music theory, required for a better understanding of our process of music composition. A comprehensive review of musical study can be found in [13].

Note is a sound with a particular pitch, duration, intensity and timbre. The pitch is the quality of sound which is

[1]http://www.jfugue.org/

recognized if it is high or low. Duration is linked to time (e.g. in seconds) that the sound vibration lasts.

Intensity, this quality determines whether the sounds are loud or soft. The timbre is the perceived sound quality of a musical note, sound or tone. We use the traditional twelve notes, *i.e.*, natural notes (C, D, E, F, G, A, B) and accidental notes (C#/Db, D#/Eb, F#/Gb, G#/Ab, A#/Bb).

Octave is the interval between one musical pitch and another with double its frequency. When we play the first twelve notes we can say that we use the first octave, if we play the next twelve, we will have played the second octave and so on. The first octaves make more bass sounds, while the latter sounds more acute. The most acute octaves cause a feeling of positivism, while more bass octaves a sense of negativity.

Scale is any set of musical notes ordered by fundamental frequency or pitch. An ascending scale is ordered by increasing pitch and a descending scale by decreasing pitch [14]. Literature confirms that the major scales reflect a feeling of *joy* and the minor scales a feeling of *sadness* or *disgust* [15].

Harmony is the way to combine sounds simultaneously. Those harmonies that are consonant, pleasant for the listener, are associated with expressions of *happiness*, *serenity*, *dreamers*, and those harmonies that are more complex and dissonant, express emotion, *tension*, *anger* and *nonconformity*.

Tempo is the speed with which the piece is played. Most used *tempo* – and its average value of beats per minute (bpm) – for composing are: *Adagio* (66-76bpm), *Moderato* (108-120bpm), *Allegro* (120-168bpm), *Presto* (168-200bpm), *Prestissimo* (200-208bpm). A fast tempo is often associated with expressions of emotion, *happiness*, *joy*, *pleasure*, *surprise*, *anger*, *anxiety*, *fear*, emotions that manifest themselves in a particular context as strong emotions. While a slow tempo is associated with expressions of *calm*, *serenity*, *sadness* [15].

Melody is the way to combine sounds as a linear succession of musical notes. Hence, many instruments are called melodic, for example, a flute, a sax, a clarinet or any wind instrument, because they can not produce more than one note at a time.

B. Extracting emotions and characters from literary works

We start our pipeline by text processing of literary works. Our methodology allows us to summarize any literary work in a PDF file format by performing the following procedure:

1) Preprocessing: It consists of discarding the information that is not relevant within the text, and that is going to be introduced to the model. To do so the natural language toolkit NLTK[2] is used. We perform a set of traditional operations, *i.e.*, conversion of abbreviations and contractions, stopwords and punctuation symbols removal and stemming. In addition, a removal of headers, table of contents and glossaries must be performed and it will be necessary to homogenize all quotation marks into double quotation marks to denote dialogues.

2) Counting emotion occurrences: We make use of a lexicon of emotions. We choose NRC Word-Emotion Association Lexicon[3] due to its large number of words associated to emotions. NRC gives us the eight basic emotions proposed by Robert Plutchik [16], of which we will use six of them - *i.e.*, *joy*, *sadness*, *anger*, *fear*, *surprise*, and *anticipation* - detailed in the Table II. For our purposes, four of these emotions were categorized as primary and the other two as secondary. In this way, *joy*, *sadness*, *anger* and *fear* are primary emotions and *surprise* and *anticipation* are secondary emotions. Then, with the text already preprocessed, a count is made of the basic emotions, in addition to the two feelings (positive or negative) that the text reflects.

3) Main characters and dialogues identification: Our preprocessing step extracts both characters and dialogues to be used by our visual representation, detailed in Section III-D. The Spacy[4] library is used to identify the characters of the literary work. Then, we develop an algorithm that allows us to group the co-references of the characters, *e.g.* Wendy is the same as Wendy Darling and Miss. Darling. Thus, in the graph only one of the co-references is plotted. Next, we identify the dialogues of the literary work by recognizing the text within each pair of double quotation marks. Finally, the dialogs that correspond to the same conversation in the novel must be grouped together. A comprehensive review of this method can be found in [7].

C. Composing music by emotions

The second step of our pipeline focus on a musical composition based on emotions. It includes three main modules, necessary to compose a melody with multiple instruments, *i.e.*, harmony, rhythm, and melody. We use the information received in the previous step, *i.e.*, the number of words associated with a primary emotion – as *joy*, *sadness*, *anger*, *fear* –, the number of words associated with a secondary emotion – as *surprise* or *anticipation* –, and the number of associated words with a positive and negative feeling. In addition, the total number of words is computed (without considering those that were eliminated by being stopwords). We split the entire literary work into sections, so this count is performed for each section (S_i) of the novel. In this work, we use four sections for processing each novel however any number greater than zero can be used.

1) Harmony module: The aim is to choose a progression of chords that best fits each emotion. This progression allows us to generate tension or resolution in a melody. We rely on the most popular chords described in Figure 2. Depending on the order in which these chords are available, it is possible to generate a lot or little tension. Thus, the chords that are closest to the beginning of the curve (I, vi, i, VI), are those that generate less tension, and the farther the next chord is chosen with respect to the beginning of the curve, the more tension there will be (V, V7).

To assign the progression of chords for each S_i, we find the distance of the most predominant emotion in S_i to the most predominant emotion in S_{i+1}. Three lines are formed with positive, negative or zero slopes, as shown in Figure 3. If the

[2]http://www.nltk.org/

[3]http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

[4]https://spacy.io/

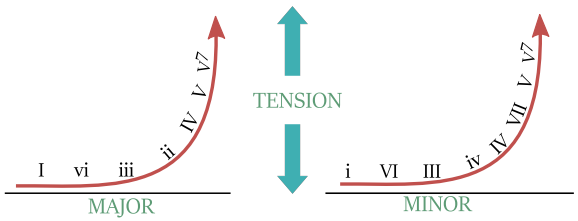


Fig. 2. Most popular chords for major and minor scales according to [17].

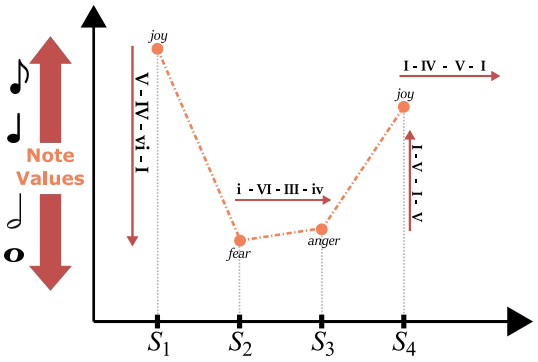


Fig. 3. Chord progressions selected for S_1 of *"Peter Pan and Wendy"*.

slope is negative or equal to zero, we will choose a progression of chords that generates high tension, and then we will add other chords that decrease the tension. Otherwise, if the slope is positive, it will start with low tension chords progressing towards a high tension.

This process will be done for the first three sections of the novel and only in the last section (S_4), a classical chord progression will be used as (I-IV-V-I). Once a chord progression has been chosen for each section, some color will be added to the notes to accentuate the emotions. For example, if the predominant emotion in S_i is *joy*, you can choose to augment the major chords and you should choose as many major chords as possible from the scale. An example of this can be found in Figure 3.

2) Rhythm module: We decompose it into three main steps. *First*, we must select a value of *Tempo*. For that, we need to know how calming or exciting the content of the literary work is. We will use the formula of Davis and Mohammad [1] with the difference that the tempo will vary between 60 and 180 bpm. The term Act is the difference between active and passive emotions densities of the whole text, where *joy* and *anger* are considered active emotions, in contrast to *sadness* considered a passive emotion. We refer by density to the count of a specific emotion divided by the total number of words. In addition, the Act value was calculated for a entire collection of literary works where the minimum and maximum values are represented by Act_{min} and Act_{max} in Equation 1.

$$tempo = 60 + \frac{(Act - Act_{min}) * (180 - 60)}{Act_{max} - Act_{min}} \quad (1)$$

Second, we choose the *Time Signature*. In our model, we limited it to a 4/4 compass which is the simplest in music.

Third, we generate a rhythm pattern. To do this, we will again refer to Figure 3. We use four possible values (whole note, half note, quarter note, eighth note). The longer the distance from one section to the other, the greater variety of note values is covered. In Figure 3, the first distance (S_1 to S_2) starts from the highest point to the lowest point of the work, so the rhythm may be the result of combining shorter duration notes (eighth notes) and longer duration notes (whole notes). While in the second distance (S_2 to S_3), only whole notes will be available due it represents the shortest distance of the work and covers less lengths. The distribution of the note values is computed dividing the distance from the highest to the lowest point by the number of sections, and locating the durations between sectors. In addition, if the slope is positive, the notes will be ordered in an increasing way, otherwise they will be arranged in a decreasing order, reinforcing the sense of direction of the rhythm. Finally, our rhythmic pattern is composed of four main measures, so its repetition aims to sound less cohesive. Thus, a total of eight measures per section are produced.

3) Melody module: Our last module consists in scale and octave selection and assign music notes to our rhythmic pattern. The scale is chosen according to the most predominant emotion in each S_i. It is important to consider that if the emotion with higher predominance belongs to one of the primary emotions (*joy*, *sadness*, *anger*, *fear*), the scale selection will be made based on this emotion. However, if the emotion that predominates is a secondary emotion (*surprise* or *anticipation*), it will be necessary to find the primary emotion that predominates in the novel (*eN*), where for those cases, the scale will be selected based on both emotions. In Table I we attach a set of scales that fits each emotion. The parameter eS refers to the most predominant emotion in S_i.

TABLE I
SET OF SCALES ASSOCIATED WITH EACH EMOTION BASED ON [18] [19]

	Emotion	**Scale**
eS = *Anticipation*	eN = *joy*	Dmaj, F#maj, Bbmaj
	eN = *sadness*	Dmin, F#min, Bbmin
	eN = *fear*	D#min , Bbmin
	eN = *anger*	Fmaj, Abmaj
eS = *Surprise*	eN = *joy*	Amaj, Bbmaj
	eN = *sadness*	Emin, Fmin
	eN = *fear*	Fmin, Bbmin, C#min
	eN = *anger*	Gmin, Bmaj
eS if *otherwise*	eS = *joy*	Cmaj, Gmaj
	eS = *sadness*	Cmin, C#min, Bmin
	eS = *fear*	D#min, Fmin, C#min
	eS = *anger*	Abmin, Fmaj, F#min

Then, we need to define in which octave of the piano the melody will be played. Thus, we use the following calculation proposed by [1]:

$$Octave = 4 + \frac{(JS - JS_{min}) * (6 - 4)}{JS_{max} - JS_{min}}, \quad (2)$$

where JS denotes the difference between the densities of *joy* and *sadness* in each section S_i, and JS_{min} and JS_{max} are the minimum and maximum value of JS from the entire collection of literary works processed, respectively.

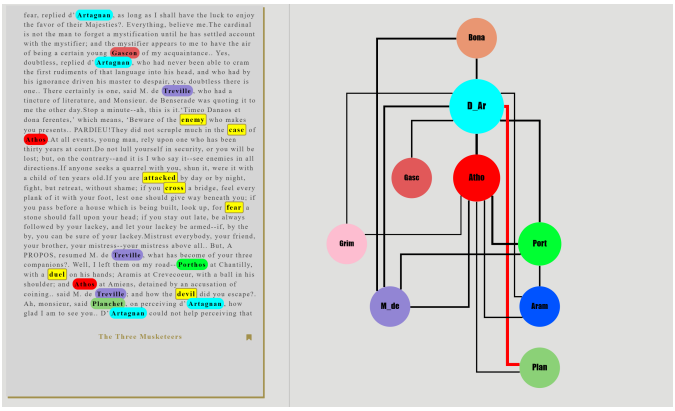


Fig. 4. An overview of our prototype: (left) an highlighted version of text dialogues and, (right) an animated graph that illustrates interaction between characters

Following, using the predominant emotion eS and the number of positive and negative words (pW, nW), we employ Equation 3 to update the $Octave$ value based on the specific emotion in Section S_i.

$$Octave = \begin{cases} Octave + 1, & \text{if } eS = joy \\ Octave - 1, & \text{if } eS = anger \\ Octave + 1, & \text{if } eS = (fear \text{ or } sadness) \\ & \text{and } pW > nW \\ Octave - 1, & \text{if } eS = (fear \text{ or } sadness) \\ & \text{and } pW < nW \\ Octave, & \text{otherwise} \end{cases} \quad (3)$$

Our next step is to assign notes into our rhythmic pattern. The selection of the notes must allow the melody to sound consonant for emotions as *joy* or *sadness* and dissonant for emotions such as *fear* or *anger*. Therefore to produce a greater consonance, we choose most of the notes into the selected chord and the remaining notes into the selected scale. In contrast, we produce dissonance (tension) by choosing notes that do not belong to the same chord or scale.

Additionally, some ornaments were added to build the melody as flourish or notes of passage, offering a greater melodic variety. Finally, to accentuate the predominant emotion in each melody and add color, a list of instruments of the JFugue were used, detailed in Table II. As a result, the .midi file generated by the JFugue plays the melody under the structure of a musical *fugue*[5]. For this reason, the chosen instruments were added and removed along the whole melody. In some cases, we generated a second melody based on the second most predominant emotion in each S_i and play it in parallel, to enhance the quality of the composition.

D. Summarizing interactions visually by graph animation

With the purpose of complementing the set of insights provided by our methodology, we decided to introduce a visual resource. Thus, we implement a graph-based animation by representing how characters interact among them, and discovering who are the most named in the literary work, and which and with who he/she relates more.

As commented in Section III-B, our preprocessing step extracts the dialogues and characters. We used this information to display the frequency variation of each character along the story by using graph's nodes, *i.e.*, each time that a character dialogues with another the size of its node is increased. The relationships among characters are represented by node's edges, where the width of the edge represents frequency of interaction between two specific characters, *i.e.*, the greater the number of dialogues between two characters in the literary work, the greater the width of the edge that joins its two nodes.

In our implementation, we use Graphviz[6] visualization software for automatically generating of graphs due to its good performance in terms of processing time and flexibility for graph drawing and D3.js[7] for embedding graphs into web context and generating transitions. Moreover, we introduce a set of constraints in order to produce a readable and aesthetic representation of the graph – *e.g.*, overlap removal, hierarchical layout and proximity between nodes – by using DOT tool.

In Figure 4, we show an overview of our prototype. On the right side, the animated graph to illustrate the interactions between characters. On the left side, we show the dialogues in detail. Colors are preserved in both sides of prototypes to rapidly identify characters. Moreover, we highlight (in light yellow) main words/actions that dialogues tells. This simple feature helps users to perceptively infer the reason that relates two or more characters.

TABLE II
LIST OF INSTRUMENT ASSOCIATED TO EACH EMOTION BASED ON [20]

Emotion	Instruments
Joy	Piano, Guitar, Flute, Piccolo, Marimba
Sadness	Piano, Glockenspiel, Orchestral-String, Celesta
Anger	Timpani, Trombone, Viola, Cello, Contra-bass
Fear	Violin, Reverse Cymbal, Vibraphone
Surprise	Violin, Piano.
Anticipation	Piano, Flute, Vibraphone

IV. RESULTS

In this section, we present the results produced by our proposed methodology. We select a subset of five best sellers of literature – i.e. *"Peter Pan and Wendy"* by James Barrie, *"Psycho"* by Robert Bloch, *"The Lovely Bones"* by Alice Sebold, *"Don Quixote of La Mancha"* by Miguel de Cervantes, and *"The Three Musketeers"* by Alexandre Dumas – in order to assess its summarization capabilities.

Table III details the extracted features by each one of the four sections (S_1 - S_4) in each literary book, namely, Activity, JS value, e_1 and e_2 emotions and Octaves. As can be noticed by simple inspection, most of obtained values are the expected

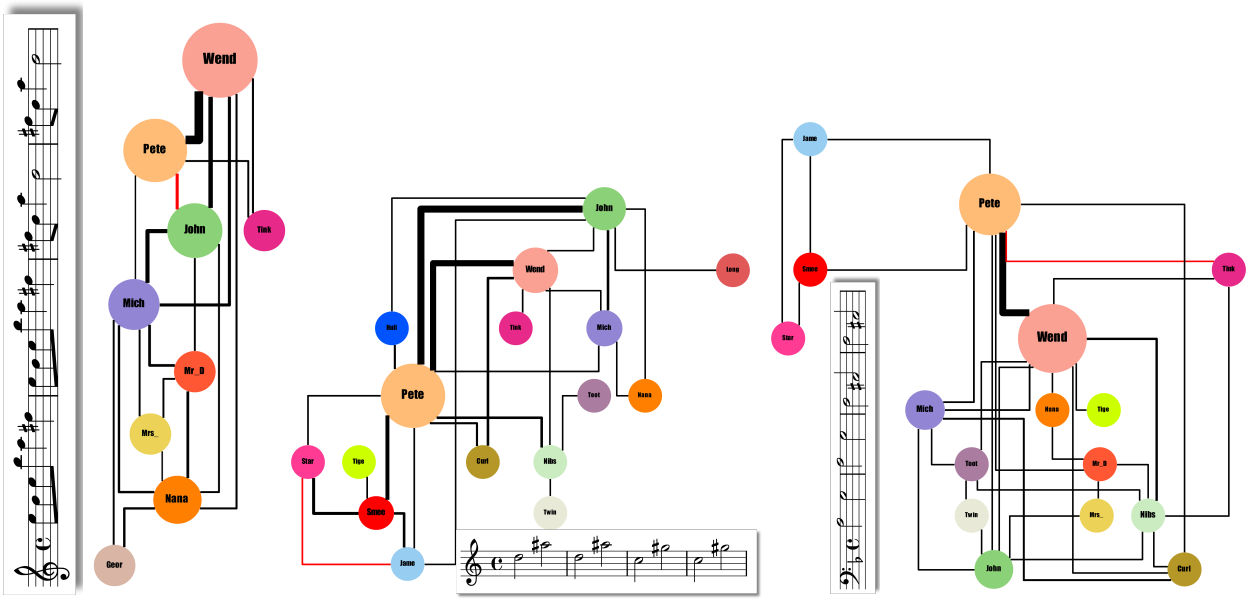
[5]A form of composition in which a theme is introduced by one voice, and imitated by other voices in succession.

[6]www.graphviz.org/

[7]www.d3js.org

TABLE III
EXTRACTED EMOTIONS, ACTIVITY, TEMPO, OCTAVES BY SECTION IN EACH PROCESSED LITERARY WORK

	"Peter Pan and Wendy"	*"Psycho"*	*"The Lovely Bones"*	*"Don Quixote of La Mancha"*	*"The Three Musketeers"*
S_1-Activity	0.058	0.029	0.032	0.057	0.043
S_1-JS	0.0313	-0.0055	0.0051	0.0129	0.0082
S_1-Tempo	180	88	98	177	120
S_1-e_1-e_2	*joy-anticipation*	*anticipation-fear*	*anticipation-joy*	*anticipation-joy*	*anticipation-joy*
S_1-Octaves	7 - 6	5 - 6	5 - 6	6-6	6-6
S_2-Activity	0.035	0.025	0.03	0.057	0.042
S_2-JS	-0.0051	-0.0056	0.005	0.0113	0.0097
S_2-Tempo	107	76	92	177	133
S_2-e_1-e_2	*fear-anger*	*anticipation-surprise*	*anticipation-sadness*	*fear-fear*	*surprise-joy*
S_2-Octaves	4 - 4	5 - 5	5 - 6	6-6	6-6
S_3-Activity	0.053	0.03	0.038	0.057	0.048
S_3-JS	0.0011	0.0004	0.0103	0.0236	0.0103
S_3-Tempo	164	92	117	177	133
S_3-e_1-e_2	*anger-joy*	*joy-fear*	*anger-surprise*	*joy-joy*	*anger-fear*
S_3-Octaves	4 - 6	6 - 6	4 - 5	6-6	4-6
S_4-Activity	0.043	0.033	0.026	0.058	0.039
S_4-JS	0.0042	0.0007	0.0048	0.0236	-0.0066
S_4-Tempo	133	94	79	180	110
S_4-e_1-e_2	*joy-anticipation*	*joy-sadness*	*sadness-surprise*	*surprise-anger*	*sadness - fear*
S_4-Octaves	6-6	6-6	6-6	5-4	6-6



(a) (b) (c)

Fig. 5. We show the resulting graphs by processing three sections of *"Peter Pan and Wendy"*. Red lines show the current interaction in the animated graph.

due to book type, *e.g.*, *joy* and *anticipation* for *"Peter Pan and Wendy"*, *anticipation* and *surprise* in *"The Lovely Bones"* and *joy* and *fear* in *"The Three Musketeers"*. Analyzing the JS value by each book, we can observe that in most of cases this value is above zero, which means that in general emotions extracted are positive, except in the specific case of *"Psycho"* that present values very close and under zero, which reflects the negative emotions identified. It also occurs in a few sections of the rest of books, specifically in S_2 of *"Peter Pan and Wendy"* and S_4 of *"The Three Musketeers"*. An interesting issue is on Activity values, since the range values (maximum $-$ minimum) among sections in the same book is about 0.01, however, *"Peter Pan and Wendy"* differs widely (0.023) presenting a high variability of emotions activity. On the other hand, *"Don Quixote of La Mancha"* (0.001) with a lowest value of variability of activity. We used all of these values to compose the music and generate the animation that conveys accurately the emotions extracted from text.

Figure 5 shows the resulting graphs for the three sections of *"Peter Pan and Wendy"*. Additionally, we show in a pentagram an initial part of the musical composition based on the extracted emotions. The leftmost graph (Figure 5a) shows a bigger interactions between Wendy and Peter comparing with other characters, where *joy* and *anticipation* predominate as

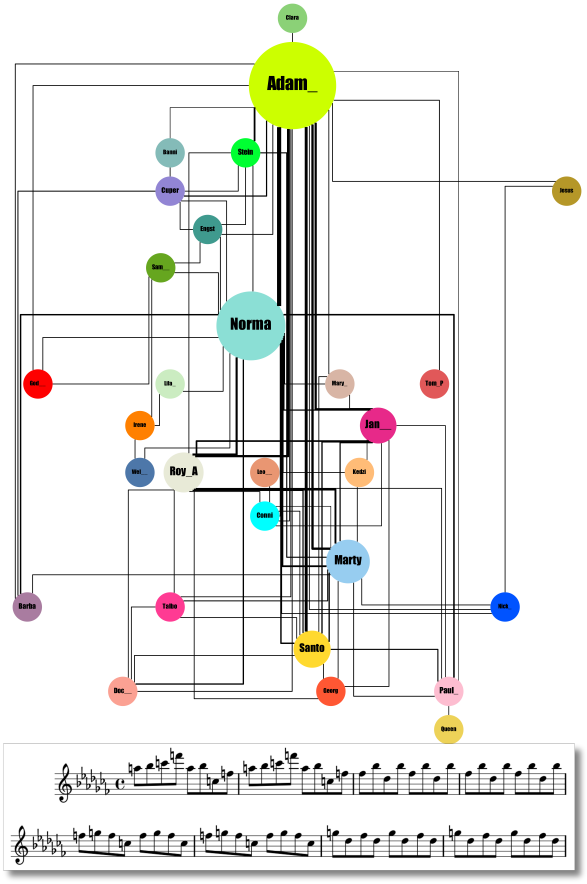


Fig. 6. Final representation by processing entirely *"Psycho"* by Robert Bloch.

main emotions. However, on the top graph (Figure 5b) which represents the second section, the biggest relationship involves to Peter and John. On the rightmost graph (Figure 5c), again the protagonism of dialogues returns to Wendy.

In Figure 6, we process entirely *"Psycho"*, *i.e.* without dividing by section, for generating a global summary of the book. As can be noticed, the highest number of dialogues involves to Adam, followed by Norma, Marty and Roy_A, respectively. This global map reveals important issues, as characters that never relate to each other (*e.g.*, as Barbara and Jesus, or Queen and Norma), or even characters that dialogues with only one other along the entire story (*e.g.*, as Clara with Adam, or Queen and Paul). About the resulting musical composition, it presents the lowest *Tempo*, causing a slow velocity comparing with the other. The predominant emotion is *fear*, so chord progressions that generate tension and a minor scale were chosen, and notes in the melody were added off the scale to add dissonance. Moreover, the choice of instruments such as Glockenspiel and Orchestral Strings accentuate the color of the emotion.

All of our results have been included in the supplementary material that supports this manuscript.

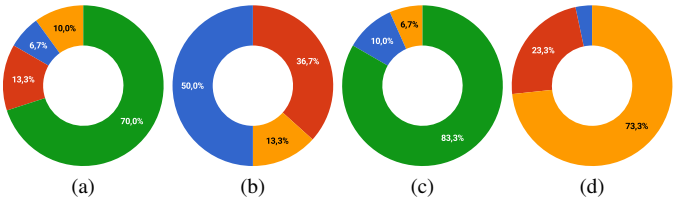


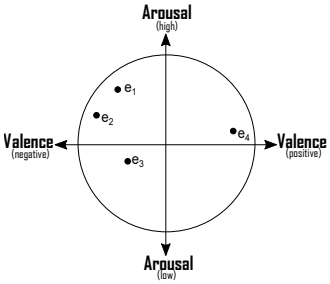
Fig. 7. Results obtained by our first user study on four different musical compositions: (a) *"Peter Pan and Wendy"* (*joy*), (b) *"Psycho"* (*fear*), (c) *"Don Quixote of La Mancha"* (*joy*), and (d) *"The three musketeers"* (*sadness*). Colors represent the chosen emotion ■ = *fear*, ■ = *angry*, ■ = *sadness*, and ■ = *joy* by each one of the 30 participants.

V. User Study

We conducted two user studies in order to assess the performance of our proposal in terms of summarization capabilities and emotion conveying. The first test validates the emotion conveyed by our musical composition. The second measures the user experience after interacting with our prototype.

A. Testing musical composition

We use four of our results (included in supplementary material) to confirm if our musical composition conveys the target emotion extracted from text, *i.e.*, *"Peter Pan and Wendy"* (*joy*), *"Psycho"* (*fear*), *"Don Quixote of La Mancha"* (*joy*), and *"The three musketeers"* (*sadness*). Each of these sections express an specific emotion described into parenthesis. We use an Arousal-Valence scheme that support us to fixed the user emotion into a 2-dimensional map, illustrated in the inline figure, where *Arousal*-axis measures how calming (low) or exciting (high) the information is, and *Valence*-axis codes emotional events as positive or negative. Then, we positioned four points as possible emotions for our experiment, *i.e.*, e1 (■) is *fear*, e2 (■) is *anger*, e3 (■) is *sadness*, and e4 (■) is *joy*. We gathered a group of 30 participants, all of them were undergraduate students (21–24 years) to play the above-mentioned four compositions. Then, we ask them to assign each one to the point that they consider it fits better. Figure 7 summarizes the produced results for each case. Note that, most of participants affirm to perceive the emotion that we intend to convey, *i.e.*, 70% *joy*, 83.3% *joy* and 73.3% *sadness*, as shown Figures 7a, 7c, and 7d respectively. However, in Figure 7b, even most of candidates (50%) confirm the correct emotion (*fear*), exist a portion (36.7%) that choose the most similar to it (*anger*). It represents a difficult decision since both emotions are placed in the same quadrant and have very close values for the arousal and valence attributes.



B. Querying on user experience by using our prototype

We accomplish this task by gathering seven participants, different ones to the previous study, to read the first section

of *"Peter Pan and Wendy"* (Included in the supplementary material) with a length of 35 pages. Then, we ask for answering the questions detailed in Table IV about their opinion on capabilities and limitations of our prototype, where the value of one ■ (1) rank as lowest and five ■ (5) as highest. Figure 8 summarizes all participant responses. As can be seen, Q1 and Q2 evaluate the musical composition and graph-based animation separately, where most participants agree that both perform satisfactorily. Although in Q3, we queried on the perception of our prototype, showing a decreasing value (three people rank as a medium the level of assimilation). This information is very useful for our purposes to keep going on research on this topic.

TABLE IV
USER EXPERIENCE QUESTIONS IN OUR SECOND USER STUDY.

ID	Question
Q1	*Does the musical composition in combination with the visualization allow you to generate a concept about the content of the section of the work?*
Q2	*Do you consider that the graph-based animation helps for improving the understanding of the interaction between the characters in the section of the work?*
Q3	*Do you consider that the prototype summarizes in an easy way to assimilate (5) or overloads your perception by the use of animation and musical composition (1)?*

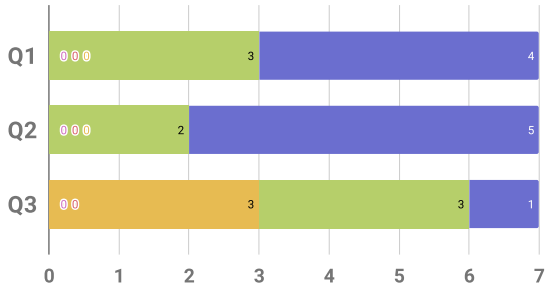


Fig. 8. Results obtained by our second user study on first section of *"Peter Pan and Wendy"*. Colors represent the chosen value ■ (1), ■ (2), ■ (3), ■ (4), and ■ (5) by each one of the 7 participants.

VI. CONCLUSION, LIMITATION AND FUTURE WORK

In this paper, we present a methodology for generating audiovisual summaries from literary works. Our approach uses emotion extraction by performing lexical analysis to generate musical compositions and graph-based animations that convey such emotions. Our prototype aims, by multimedia resources, to increase the engagement of users for discovering in-detail the story behind our summary.

Currently, there are not many related works to our proposal, making it very difficult to find information that would serve as our starting point. Especially in the part of musical composition based on emotions, not only there is a great shortage of algorithms focused on affective composition, but there is also a great absence of data set associated with emotions. On the one hand, we find very few lexicons whose data are associated with an emotion and on the other hand, machine learning methods have not reached high accuracy in their predictions, precisely because of the lack of data.

As future work, we think that implementing some interactive tools for user exploration could increase analysis capabilities to improve readability and understanding of books.

ACKNOWLEDGMENT

The authors acknowledge the financial support from Department of Computer Science at the Universidad Católica San Pablo and São Paulo Research Foundation - FAPESP (grant #2019/10560-0).

REFERENCES

[1] H. Davis and S. M. Mohammad, "Generating music from literature," *CoRR*, vol. abs/1403.2124, 2014. [Online]. Available: http://arxiv.org/abs/1403.2124
[2] D. K. Elson, N. Dames, and K. R. McKeown, "Extracting social networks from literary fiction," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ser. ACL '10. Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, pp. 138–147.
[3] D. Williams, A. Kirke, E. R. Miranda, E. Roesch, I. Daly, and S. Nasuto, "Investigating affect in algorithmic composition systems," *Psychology of Music*, vol. 43, no. 6, pp. 831–854, 2015.
[4] M. Yamada, Y. Murai, and I. Kumagai, "Story visualization of novels with multi-theme keyword density analysis," *Journal of Visualization*, vol. 16, no. 3, pp. 247–257, Aug 2013.
[5] M. Yamada and Y. Murai, "Story visualization of literary works," *Journal of Visualization*, vol. 12, no. 2, pp. 181–188, Jun 2009.
[6] S. Mohammad, "From once upon a time to happily ever after: Tracking emotions in novels and fairy tales," in *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, ser. LaTeCH '11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 105–114.
[7] M. C. Waumans, T. Nicodème, and H. Bersini, "Topology analysis of social networks extracted from literature," *PLOS ONE*, vol. 10, no. 6, pp. 1–30, 06 2015.
[8] L. Zhang and B. Liu, *Sentiment Analysis and Opinion Mining*. Boston, MA: Springer US, 2017, pp. 1152–1161.
[9] J. Tao and T. Tan, "Affective computing: A review," in *Affective Computing and Intelligent Interaction*, J. Tao, T. Tan, and R. W. Picard, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 981–995.
[10] C.-C. Stere and S. Trausan-Matu, "Generation of musical accompaniment for a poem, using artificial intelligence techniques," *Romanian Journal of Human - Computer Interaction*, vol. 10, no. 3, pp. 250–270, 2017.
[11] F. Morreale and A. De Angeli, "Collaborating with an autonomous agent to generate affective music," *Comput. Entertain.*, vol. 14, no. 3, pp. 5:1–5:21, Dec. 2016.
[12] L. Yang, S. Chou, and Y. Yang, "Midinet: A convolutional generative adversarial network for symbolic-domain music generation using 1d and 2d conditions," *CoRR*, vol. abs/1703.10847, 2017.
[13] H. Baxter and M. Baxter, *The Right Way to Read Music*, ser. Right way. Right Way, 2008.
[14] B. Benward and M. Saker, *Music in Theory and Practice, Volume 1 with Audio CD*, ser. Music in Theory and Practice. McGraw-Hill Education, 2008.
[15] A. Gabrielsson and E. Lindström, "The influence of musical structure on emotional expression." 2001.
[16] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word-emotion association lexicon," vol. 29, no. 3, pp. 436–465, 2013.
[17] F. Findeisen, *The Addiction Formula*. Albino Publishing, 2015.
[18] W. Chase, *How Music Really Works*. Roedy Black Publishing; Edition: Second Edition (2006), 2006.
[19] J. Hobbs. (2018) Musical key characteristics and emotions. [Online]. Available: https://ledgernote.com/blog/interesting/musical-key-characteristics-emotions/
[20] P. Juslin, P. N. Juslin, J. Sloboda, and P. Sloboda, *Handbook of Music and Emotion: Theory, Research, Applications*, ser. Affective Science. OUP Oxford, 2010.