

Generating Music from Literature

Hannah Davis

New York University
hannah.davis@nyu.edu

Saif M. Mohammad

National Research Council Canada
saif.mohammad@nrc-cnrc.gc.ca

Abstract

We present a system, *TransProse*, that automatically generates musical pieces from text. *TransProse* uses known relations between elements of music such as tempo and scale, and the emotions they evoke. Further, it uses a novel mechanism to determine sequences of notes that capture the emotional activity in text. The work has applications in information visualization, in creating audio-visual e-books, and in developing music apps.

1 Introduction

Music and literature have an intertwined past. It is believed that they originated together (Brown, 1970), but in time, the two have developed into separate art forms that continue to influence each other.¹ Music, just as prose, drama, and poetry, is often used to tell stories.² Opera and ballet tell stories through music and words, but even instrumental music, which is devoid of words, can have a powerful narrative form (Hatten, 1991). Mahler’s and Beethoven’s symphonies, for example, are regarded as particularly good examples of narrative and evocative music (Micznik, 2001).

In this paper, for the first time, we present a method to automatically generate music from literature. Specifically, we focus on novels and generate music that captures the change in the distribution of emotion words. We list below some of the benefits in pursuing this general line of research:

- Creating audio-visual e-books that generate music when certain pages are opened—music that accentuates the mood conveyed by the text in those pages.

¹The term *music* comes from *muses*—the nine Greek goddesses of inspiration for literature, science, and arts.

²Music is especially close to poetry as songs often tend to be poems set to music.

- Mapping pieces of literature to musical pieces according to compatibility of the flow of emotions in text with the audio characteristics of the musical piece.
- Finding songs that capture the emotions in different parts of a novel. This could be useful, for example, to allow an app to find and play songs that are compatible with the mood of the chapter being read.
- Generating music for movie scripts.
- Appropriate music can add to good visualizations to communicate information effectively, quickly, and artfully.

Example 1: A tweet stream that is accompanied by music that captures the aggregated sentiment towards an entity.

Example 2: Displaying the world map where clicking on a particular region plays music that captures the emotions of the tweets emanating from there.

Given a novel (in an electronically readable form), our system, which we call *TransProse*, generates simple piano pieces whose notes are dependent on the emotion words in the text. The challenge in composing new music, just as in creating a new story, is the infinite number of choices and possibilities. We present a number of mapping rules to determine various elements of music, such as tempo, major/minor key, etc. according to the emotion word density in the text. We introduce a novel method to determine the sequence of notes (sequences of pitch and duration pairs) to be played as per the change in emotion word density in the text. We also list some guidelines we followed to make the sequence of notes sound like music as opposed to a cacophonous cascade of sounds.

Certainly, there is no one right way of capturing the emotions in text through music, and there is no one right way to produce good music. Generating

compelling music is an art, and TransProse can be improved in a number of ways (we list several advancements in the Future Work section). Our goal with this project is to present initial ideas in an area that has not been explored before.

This paper does not assume any prior knowledge of music theory. Section 2 presents all the terminology and ideas from music theory needed to understand this paper. We present related work in Section 3. Sections 4, 5, 6, and 7 describe our system. In Sections 8 and 9, we present an analysis of the music generated by our system for various popular novels. Finally in Section 10 we present limitations and future work.

2 Music

In physical terms, music is a series of possibly overlapping sounds, often intended to be pleasing to the listener. Sound is what our ears perceive when there is a mechanical oscillation of pressure in some medium such as air or water. Thus, different sounds are associated with different frequencies. In music, a particular frequency is referred to as *pitch*. A *note* has two aspects: pitch and relative duration.³ Examples of relative duration are *whole-note*, *half-note*, *quarter-note*, etc. Each successive element in this list is of half the duration as the preceding element. Consider the example notes: 400Hz–quarter-note and 760Hz–whole-note. The first note is the sound corresponding to 400Hz, whereas the second note is the sound corresponding to 760Hz. Also, the first note is to be played for one fourth the duration the second note is played. It is worth repeating at this point that *note* and *whole-note* do not refer to the same concept—the former is a combination of pitch and relative duration, whereas *whole-note* (and others such as *quarter-note* and *half-note*) are used to specify the relative duration of a note. Notes are defined in terms of *relative* duration to allow for the same melody to be played quickly or slowly.

A series of notes can be grouped into a *measure* (also called *bar*). *Melody* (also called *tune*) is a sequence of measures (and therefore a sequence of notes) that creates the musical piece itself. For example, a melody could be defined as 620Hz–half-note, 1200Hz–whole-note, 840Hz–half-note,

³Confusingly, *note* is also commonly used to refer to *pitch* alone. To avoid misunderstanding, we will not use *note* in that sense in this paper. However, some statements, such as *play that pitch* may seem odd to those familiar with music, who may be more used to *play that note*.

660Hz–quarter-note, and so on. There can be one melody (for example, in the song *Mary Had A Little Lamb*) or multiple melodies; they can last throughout the piece or appear in specific sections. A challenge for TransProse is to generate appropriate sequences of notes, given the infinite possibilities of pitch, duration, and order of the notes.

Tempo is the speed at which the piece should be played. It is usually indicated by the number of beats per minute. A beat is a basic unit of time. A quarter-note is often used as one beat. In which case, the tempo can be understood simply as the number of quarter-notes per minute. Consider an example. Let's assume it is decided that the example melody specified in the earlier paragraph is to be played at a tempo of 120 quarter-notes per minute. The total number of quarter-notes in the initial sequence (620Hz–half-note, 1200Hz–whole-note, 840Hz–half-note, and 660Hz–quarter-note) is $2 + 4 + 2 + 1 = 9$. Thus the initial sequence must be played in $9/120$ minutes, or 4.5 seconds.

The *time signature* of a piece indicates two things: a) how many beats are in a measure, and b) which note duration represents one beat. It is written as one number stacked on another number. The upper number is the number of beats per measure, and the lower number is the note duration that represents one beat. For example, a time signature of $\frac{6}{8}$ would mean there are six beats per measure, and an eighth note represents one beat. One of the most common time signatures is $\frac{4}{4}$, and it is referred to as *common time*.

Sounds associated with frequencies that are multiples or factors of one another (for example, 440Hz, 880Hz, 1760Hz, etc) are perceived by the human ear as being consonant and pleasing. This is because the pressure waves associated with these sounds have overlapping peaks and troughs. Sets of such frequencies or pitches form pitch classes. The intervals between successive pitches in a pitch class are called *octaves*. On a modern 88-key piano, the keys are laid out in increasing order of pitch. Every successive 12 keys pertain to an octave. (Thus there are keys pertaining to 7 octaves and four additional keys pertaining to the eighth octave.) Further, each of the 12 keys split the octave such that the difference in frequency between successive keys in an octave is the same. Thus the corresponding keys in each octave form a pitch class. For example, the keys at posi-

tion 1, 13, 25, 37, and so on, form a pitch class. Similarly keys at position 2, 14, 26, 38, and so on, form another pitch class. The pitch classes on a piano are given names C, C#, D, D#, E, F, F#, G, G#, A, A#, B. (The # is pronounced *sharp*). The same names can also be used to refer to a particular key in an octave. (In an octave, there exists only one C, only one D#, and so on.) The octaves are often referred to by a number. On a standard piano, the octaves in increasing order are 0, 1, 2, and so on. C2 refers to the key in octave 2 that is in the pitch class C.⁴

The difference in frequency between successive piano keys is called a *semitone* or *Half-Tone* (*Half* for short). The interval between two keys separated by exactly one key is called *Whole-Tone* (*Whole* for short). Thus, the interval between C and C# is half, whereas the interval between C and D is whole. A *scale* is any sequence of pitches ordered by frequency. A *major scale* is a sequence of pitches obtained by applying the ascending pattern: Whole–Whole–Half–Whole–Whole–Whole–Half. For example, if one starts with C, then the corresponding C major scale consists of C, D (frequency of C + Whole interval), E (frequency of D + Whole interval), F (frequency of E + Half interval), G, A, B, C. Major scales can begin with any pitch (not just C), and that pitch is called the *base pitch*. A *major key* is the set of pitches corresponding to the major scale. Playing in the key of C major means that one is primarily playing the keys (pitches) from the corresponding scale, C major scale (although not necessarily in a particular order).

Minor scales are series of pitches obtained by applying the ascending pattern: Whole–Half–Whole–Whole–Half–Whole–Whole. Thus, C minor is C, D, D#, F, G, G#, A#, C. A *minor key* is the set of pitches corresponding to the minor scale. Playing in major keys generally creates lighter sounding pieces, whereas playing in minor keys creates darker sounding pieces.

Consonance is how pleasant or stable one perceives two pitches played simultaneously (or one after the other). There are many theories on what makes two pitches consonant, some of which are

culturally dependent. The most common notion (attributed to Pythagoras) is that the simpler the ratio between the two frequencies, the more consonant they are (Roederer, 2008; Tenney, 1988).

Given a particular scale, some have argued that the order of the pitches in decreasing consonance is as follows: 1st, 5th, 3rd, 6th, 2nd, 4th, and 7th (Perricone, 2000). Thus for the C major—C (the base pitch, or 1st), D (2nd), E (3rd), F (4th), G (5th), A (6th), B (7th)—the order of the pitches in decreasing consonance is—C, G, E, A, D, F, B. Similarly, for C minor—C (the base pitch, or 1st), D (2nd), D# (3rd), F (4th), G (5th), G# (6th), A# (7th)—the order of pitches in decreasing consonance is—C, G, D#, G#, D, F, A#. We will use these orders in TransProse to generate more discordant and unstable pitches to reflect higher emotion word densities in the novels.

3 Related Work

This work is related to automatic sentiment and emotion analysis of text (computational linguistics), the generation of music (music theory), as well as the perception of music (psychology).

Sentiment analysis techniques aim to determine the evaluative nature of text—*positive*, *negative*, or *neutral*. They have been applied to many different kinds of texts including customer reviews (Pang and Lee, 2008), newspaper headlines (Bellegarda, 2010), emails (Liu et al., 2003; Mohammad and Yang, 2011), blogs (Genereux and Evans, 2006; Mihalcea and Liu, 2006), and tweets (Pak and Paroubek, 2010; Agarwal et al., 2011; Thelwall et al., 2011; Brody and Diakopoulos, 2011; Aisopos et al., 2012; Bakliwal et al., 2012). Surveys by Pang and Lee (2008) and Liu and Zhang (2012) give a summary of many of these approaches. Emotion analysis and affective computing involve the detection of emotions such as *joy*, *anger*, *sadness*, and *anticipation* in text. A number of approaches for emotion analysis have been proposed in recent years (Boucouvalas, 2002; Zhe and Boucouvalas, 2002; Aman and Szpakowicz, 2007; Neviarouskaya et al., 2009; Kim et al., 2009; Bollen et al., 2009; Tumasjan et al., 2010). Text-to-speech synthesis employs emotion detection to produce speech consistent with the emotions in the text (Iida et al., 2000; Pierre-Yves, 2003; Schröder, 2009). See surveys by Picard (2000) and Tao and Tan (2005) for a broader review of the research in this area.

⁴The frequencies of piano keys at a given position across octaves is in log scale. For example, frequencies of C1, C2, . . . , and so on are in log scale. The perception of sound (frequency) in the human ear is also roughly logarithmic. Also, the frequency 440Hz (mentioned above) is A4 and it is the customary tuning standard for musical pitch.

Some prior empirical sentiment analysis work focuses specifically on literary texts. Alm and Sproat (2005) analyzed twenty two Brothers Grimm fairy tales to show that fairy tales often began with a neutral sentence and ended with a happy sentence. Mohammad (2012) visualized the emotion word densities in novels and fairy tales. Volkova et al. (2010) study human annotation of emotions in fairy tales. However, there exists no work connecting automatic detection of sentiment with the automatic generation of music.

Methods for both sentiment and emotion analysis often rely on lexicons of words associated with various affect categories such as positive and negative sentiment, and emotions such as joy, sadness, fear, and anger. The WordNet Affect Lexicon (WAL) (Strapparava and Valitutti, 2004) has a few hundred words annotated with associations to a number of affect categories including the six Ekman emotions (joy, sadness, anger, fear, disgust, and surprise).⁵ The NRC Emotion Lexicon, compiled by Mohammad and Turney (2010; 2013), has annotations for about 14000 words with eight emotions (six of Ekman, trust, and anticipation).⁶ We use this lexicon in our project.

Automatic or semi-automatic generation of music through computer algorithms was first popularized by Brian Eno (who coined the term *generative music*) and David Cope (Cope, 1996). Lerdahl and Jackendoff (1983) authored a seminal book on the generative theory of music. Their work greatly influenced future work in automatic generation of music such as that of Collins (2008) and Biles (1994). However, these pieces did not attempt to explicitly capture emotions.

Dowling and Harwood (1986) showed that vast amounts of information are processed when listening to music, and that the most expressive quality that one perceives is emotion. The communication of emotions in non-verbal utterances and in music show how emotions in music have an evolutionary basis (Rousseau, 2009; Spencer, 1857; Juslin and Laukka, 2003). There are many known associations between music and emotions:

- *Loudness*: Loud music is associated with intensity, power, and anger, whereas soft music is associated with sadness or fear (Gabrielsson and Lindström, 2001).

⁵<http://wndomains.fbk.eu/wnaffect.html>

⁶<http://www.purl.org/net/NRCemotionlexicon>

- *Melody*: A sequence of consonant notes is associated with joy and calm, whereas a sequence of disconsonant notes is associated with excitement, anger, or unpleasantness (Gabrielsson and Lindström, 2001).
- *Major and Minor Keys*: Major keys are associated with happiness, whereas minor keys are associated with sadness (Hunter et al., 2010; Hunter et al., 2008; Ali and Peynirciolu, 2010; Gabrielsson and Lindström, 2001; Webster and Weir, 2005).
- *Tempo*: Fast tempo is associated with happiness or excitement (Hunter et al., 2010; Hunter et al., 2008; Ali and Peynirciolu, 2010; Gabrielsson and Lindström, 2001; Webster and Weir, 2005).

Studies have shown that even though many of the associations mentioned above are largely universal, one's own culture also influences the perception of music (Morrison and Demorest, 2009; Balkwill and Thompson, 1999).

4 Our System: TransProse

Our system, which we call TransProse, generates music according to the use of emotion words in a given novel. It does so in three steps: First, it analyzes the input text and generates an emotion profile. The emotion profile is simply a collection of various statistics about the presence of emotion words in the text. Second, based on the emotion profile of the text, the system generates values for tempo, scale, octave, notes, and the sequence of notes for multiple melodies. Finally, these values are provided to JFugue, an open-source Java API for programming music, that generates the appropriate audio file. In the sections ahead, we describe the three steps in more detail.

5 Calculating Emotion Word Densities

Given a novel in electronic form, we use the NRC Emotion Lexicon (Mohammad and Turney, 2010; Mohammad and Turney, 2013) to identify the number of words in each chapter that are associated with an affect category. We generate counts for eight emotions (anticipation, anger, joy, fear, disgust, sadness, surprise, and trust) as well as for positive and negative sentiment. We partition the novel into four sections representing the beginning, early middle, late middle, and end. Each section is further partitioned into four sub-sections.

The number of sections, the number of subsections per section, and the number of notes generated for each of the subsections together determine the total number of notes generated for the novel. Even though we set the number of sections and number of sub-sections to four each, these settings can be varied, especially for significantly longer or shorter pieces of text.

For each section and for each sub-section the ratio of emotion words to the total number of words is calculated. We will refer to this ratio as the *overall emotions density*. We also calculate densities of particular emotions, for example, the *joy density*, *anger density*, etc. As described in the section ahead, the emotion densities are used to generate sequences of notes for each of the subsections.

6 Generating Music Specifications

Each of the pieces presented in this paper are for the piano with three simultaneous, but different, melodies coming together to form the musical piece. Two melodies sounded too thin (simple), and four or more melodies sounded less cohesive.

6.1 Major and Minor Keys

Major keys generally create a more positive atmosphere in musical pieces, whereas minor keys tend to produce pieces with more negative undertones (Hunter et al., 2010; Hunter et al., 2008; Ali and Peynirciolu, 2010; Gabrielsson and Lindström, 2001; Webster and Weir, 2005). No consensus has been reached on whether particular keys themselves (for example, A minor vs E minor) evoke different emotions, and if so, what emotions are evoked by which keys. For this reason, the prototype of Transpose does not consider different keys; the chosen key for the produced musical pieces is limited to either C major or C minor. (C major was chosen because it is a popular choice when teaching people music. It is simple because it does not have any sharps. C minor was chosen as it is the minor counterpart of C major.)

Whether the piece is major or minor is determined by the ratio of the number of positive words to the number of negative words in the entire novel. If the ratio is higher than 1, C major is used, that is, only pitches pertaining to C major are played. If the ratio is 1 or lower, C minor is used.

Experimenting with keys other than C major and C minor is of interest for future work. Further-

more, the eventual intent is to include mid-piece key changes for added effect. For example, changing the key from C major to A minor when the plot suddenly turns sad. The process of changing key is called modulation. Certain transitions such as moving from C major to A minor are commonly used and musically interesting.

6.2 Melodies

We use three melodies to capture the change in emotion word usage in the text. The notes in one melody are based on the overall emotion word density (the emotion words associated with any of the eight emotions in the NRC Emotion Lexicon). We will refer to this melody, which is intended to capture the overarching emotional movement, as *melody o* or M_o (the ‘o’ stands for overall emotion). The notes in the two other melodies, *melody e1* (M_{e1}) and *melody e2* (M_{e2}), are determined by the most prevalent and second most prevalent emotions in the text, respectively. Precisely how the notes are determined is described in the next sub-section, but first we describe how the octaves of the notes is determined.

The octave of melody o is proportional to the difference between the joy and sadness densities of the novel. We will refer to this difference by JS . We calculated the lowest density difference (JS_{min}) and highest JS score (JS_{max}) in a collection of novels. For a novel with density difference, JS , the score is linearly mapped to octave 4, 5, or 6 of a standard piano, as per the formula shown below:

$$Oct(M_o) = 4 + r\left(\frac{(JS - JS_{min}) * (6 - 4)}{JS_{max} - JS_{min}}\right) \quad (1)$$

The function r rounds the expression to the closest integer. Thus scores closer to JS_{min} are mapped to octave 4, scores closer to JS_{max} are mapped to octave 6, and those in the middle are mapped to octave 5.

The octave of M_{e1} is calculated as follows:

$$Oct(M_{e1}) = \begin{cases} Oct(M_o) + 1, & \text{if } e1 \text{ is joy or trust} \\ Oct(M_o) - 1, & \text{if } e1 \text{ is anger, fear,} \\ & \text{sadness, or disgust} \\ Oct(M_o), & \text{otherwise} \end{cases} \quad (2)$$

That is, M_{e1} is set to:

- an octave higher than the octave of M_o if $e1$ is a positive emotion,

- an octave lower than the octave of M_o if e_1 is a negative emotion,
- the same octave as that of M_o if e_1 is surprise or anticipation.

Recall that higher octaves evoke a sense of positivity, whereas lower octaves evoke a sense of negativity. The octave of M_{e2} is calculated exactly as that of M_{e1} , except that it is based on the second most prevalent emotion (and not the most prevalent emotion) in the text.

6.3 Structure and Notes

As mentioned earlier, TransProse generates three melodies that together make the musical piece for a novel. The method for generating each melody is the same, with the exception that the three melodies (M_o , M_{e1} , and M_{e2}) are based on the overall emotion density, predominant emotion’s density, and second most dominant emotion’s density, respectively. We describe below the method common for each melody, and use *emotion word density* as a stand in for the appropriate density.

Each melody is made up of four sections, representing four sections of the novel (the beginning, early middle, late middle, and end). In turn, each section is represented by four measures. Thus each measure corresponds to a quarter of a section (a sub-section). A measure, as defined earlier, is a series of notes. The number of notes, the pitch of each note, and the relative duration of each note are determined such that they reflect the emotion word densities in the corresponding part of the novel.

Number of Notes: In our implementation, we decided to contain the possible note durations to whole notes, half notes, quarter notes, eighth notes, and sixteenth notes. A relatively high emotion density is represented by many notes, whereas a relatively low emotion density is represented by fewer notes. We first split the interval between the maximum and minimum emotion density for the novel into five equal parts (five being the number of note duration choices – whole, half, quarter, eighth, or sixteenth). Emotion densities that fall in the lowest interval are mapped to a single whole note. Emotion densities in the next interval are mapped to two half-notes. The next interval is mapped to four quarter-notes. And so on, until the densities in the last interval are mapped to sixteen sixteenth-notes ($1/16^{\text{th}}$). The result is shorter notes during periods of higher emotional activity

(with shorter notes making the piece sound more active), and longer notes during periods of lower emotional activity.

Pitch: If the number of notes for a measure is n , then the corresponding sub-section is partitioned into n equal parts and the pitch for each note is based on the emotion density of the corresponding sub-section. Lower emotion densities are mapped to more consonant pitches in the key (C major or C minor), whereas higher emotion densities are mapped to less consonant pitches in the same scale. For example, if the melody is in the key of C major, then the lowest to highest emotion densities are mapped linearly to the pitches C, G, E, A, D, F, B. Thus, a low emotion value would create a pitch that is more consonant and a high emotion value would create a pitch that is more dissonant (more interesting and unusual).

Repetition: Once the four measures of a section are played, the same four measures are repeated in order to create a more structured and melodic feeling. Without the repetition, the piece sounds less cohesive.

6.4 Tempo

We use a $\frac{4}{4}$ time signature (common time) because it is one of the most popular time signatures. Thus each measure (sub-section) has 4 beats. We determined tempo (beats per minute) by first determining how active the target novel is. Each of the eight basic emotions is assigned to be either active, passive, or neutral. In TransProse, the tempo is proportional to the *activity score*, which we define to be the difference between the average density of the active emotions (anger and joy) and the average density of the passive emotions (sadness). The other five emotions (anticipation, disgust, fear, surprise, and trust) were considered ambiguous or neutral, and did not influence the tempo.

We subjectively identified upper and lower bounds for the possible tempo values to be 180 and 40 beats/minute, respectively. We determined activity scores for a collection of novels, and identified the highest activity score (Act_{max}) and the lowest activity score (Act_{min}). For a novel whose activity score was Act , we determined tempo as per the formula shown below:

$$tempo = 40 + \frac{(Act - Act_{min}) * (180 - 40)}{Act_{max} - Act_{min}} \quad (3)$$

Thus, high activity scores were represented by

tempo values closer to 180 and lower activity scores were represented by tempo values closer to 40. The lowest activity score in our collection of texts, Act_{min} , was -0.002 whereas the highest activity score, Act_{max} , was 0.017.

7 Converting Specifications to Music

JFugue is an open-source Java API that helps create generative music.⁷ It allows the user to easily experiment with different notes, instruments, octaves, note durations, etc within a Java program. JFugue requires a line of specifically-formatted text that describes the melodies in order to play them. The initial portion of the string of JFugue tokens for the novel *Peter Pan* is shown below. The string conveys the overall information of the piece as well as the first eight measures (or one section) for each of the three melodies (or voices).

```
KCmaj X[VOLUME]=16383 V0 T180
A6/0.25 D6/0.125 F6/0.25 B6/0.25
B6/0.125 B6/0.25 B6/0.25...
```

K stands for key and *Cmaj* stands for C major. This indicates that the rest of the piece will be in the key of C major. The second token controls the volume, which in this example is at the loudest value (16383). *V0* stands for the first melody (or voice). The tokens with the letter *T* indicate the tempo, which in the case of this example is 180 beats per minute.

The tokens that follow indicate the notes of the melody. The letter is the pitch class of the note, and the number immediately following it is the octave. The number following the slash character indicates the duration of the note. (0.125 is an eighth-note (1/8th), 0.25 is a quarter note, 0.5 is a half note, and 1.0 is a whole note.) We used JFugue to convert the specifications of the melodies into music. JFugue saves the pieces as a midi files, which we converted to MP3 format.⁸

8 Case Studies

We created musical pieces for several popular novels through TransProse. These pieces are available at: <http://transprose.weebly.com/final-pieces.html>. Since these novels are

⁷<http://www.jfugue.org>

⁸The MP3 format uses a lossy data compression, but the resulting files are significantly smaller in size. Further, a wider array of music players support the MP3 format.

likely to have been read by many people, the readers can compare their understanding of the story with the music generated by TransProse. Table 1 presents details of some of these novels.

8.1 Overall Tone

TransProse captures the overall positive or negative tone of the novel by assigning an either major or minor key to the piece. *Peter Pan* and *Anne of Green Gables*, novels with overall happy and uplifting moods, created pieces in the major key. On the other hand, novels such as *Heart of Darkness*, *A Clockwork Orange*, and *The Road*, with dark themes, created pieces in the minor key. The effect of this is pieces that from the start have a mood that aligns with the basic mood of the novel they are based on.

8.2 Overall Happiness and Sadness

The densities of happiness and sadness in a novel are represented in the baseline octave of a piece. This representation instantly conveys whether the novel has a markedly happy or sad mood. The overall high happiness densities in *Peter Pan* and *Anne of Green Gables* create pieces in an octave above the average, resulting in higher tones and a lighter overall mood. Similarly, the overall high sadness densities in *The Road* and *Heart of Darkness* result in pieces an octave lower than the average, and a darker overall tone to the music. Novels, such as *A Clockwork Orange*, and *The Little Prince*, where happiness and sadness are not dramatically higher or lower than the average novel remain at the average octave, allowing for the creation of a more nuanced piece.

8.3 Activeness of the Novel

Novels with lots of active emotion words, such as *Peter Pan*, *Anne of Green Gables*, *Lord of the Flies*, and *A Clockwork Orange*, generate fast-paced pieces with tempos over 170 beats per minute. On the other hand, *The Road*, which has relatively few active emotion words is rather slow (a tempo of 42 beats per minute).

8.4 Primary Emotions

The top two emotions of a novel inform two of the three melodies in a piece (M_{e1} and M_{e2}). Recall that if the melody is based on a positive emotion, it will be an octave higher than the octave of M_o , and if it is based on a negative emotion, it will be an octave lower. For novels where the top two emotions

Table 1: Emotion and audio features of a few popular novels that were processed by TransProse. The musical pieces are available at: <http://transprose.weebly.com/final-pieces.html>.

Book Title	Emotion 1	Emotion 2	Octave	Tempo	Pos/Neg	Key	Activity	Joy-Sad
<i>A Clockwork Orange</i>	Fear	Sadness	5	171	Negative	C Minor	0.009	-0.0007
<i>Alice in Wonderland</i>	Trust	Fear	5	150	Positive	C Major	0.007	-0.0002
<i>Anne of Green Gables</i>	Joy	Trust	6	180	Positive	C Major	0.010	0.0080
<i>Heart of Darkness</i>	Fear	Sadness	4	122	Negative	C Minor	0.005	-0.0060
<i>Little Prince, The</i>	Trust	Joy	5	133	Positive	C Major	0.006	0.0028
<i>Lord of The Flies</i>	Fear	Sadness	4	151	Negative	C Minor	0.008	-0.0053
<i>Peter Pan</i>	Trust	Joy	6	180	Positive	C Major	0.010	0.0040
<i>Road, The</i>	Sadness	Fear	4	42	Negative	C Minor	-0.002	-0.0080
<i>To Kill a Mockingbird</i>	Trust	Fear	5	132	Positive	C Major	0.006	-0.0013

are both positive, such as *Anne of Green Gables* (trust and joy), the pieces sound especially light and joyful. For novels where the top two emotions are both negative, such as *The Road* (sadness and fear), the pieces sound especially dark.

8.5 Emotional Activity

Unlike the overall pace of the novel, individual segments of activity were also identified in the pieces through the number and duration of notes (with more and shorter notes indicating higher emotion densities). This can be especially heard in the third section of *A Clockwork Orange*, the final section of *The Adventures of Sherlock Holmes*, the second section of *To Kill a Mockingbird*, and the final section of *Lord of the Flies*. In *A Clockwork Orange*, the main portion of the piece is chaotic and eventful, likely as the main characters cause havoc; at the end of the novel (as the main character undergoes therapy) the piece dramatically changes and becomes structured. Similarly, in *Heart of Darkness*, the piece starts out only playing a few notes; as the tension in the novel builds, the number of notes increases and their durations decrease.

9 Comparing Alternative Choices

We examine choices made in TransProse by comparing musical pieces generated with different alternatives. These audio clips are available here: <http://transprose.weebly.com/clips.html>.

Pieces with two melodies (based on overall emotion density and the predominant emotion’s density) and pieces based on four melodies (based on the top three emotions and the overall emotion density) were generated and uploaded in the clips webpage. Observe that with only two melodies, the pieces tend to sound thin, whereas with four melodies the pieces sound less cohesive and some-

times chaotic. The effect of increasing and decreasing the total number of sections and sub-sections is also presented. Additionally, the webpage displays pieces with tempos and octaves beyond the limits chosen in TransProse. We also show other variations such as pieces for relatively positive novels generated in C minor (instead of C major). These alternatives are not necessarily incorrect, but they tend to often be less effective.

10 Limitations and Future work

We presented a system, TransProse, that generates music according to the use of emotion words in a given piece of text. A number of avenues for future work exist such as exploring the use of mid-piece key changes and intentional harmony and discord between the melodies. We will further explore ways to capture activity in music. For example, an automatically generated activity lexicon (built using the method proposed by Turney and Littman (2003)) can be used to identify portions of text where the characters are relatively active (fighting, dancing, conspiring, etc) and areas where they are relatively passive (calm, incapacitated, sad, etc). One can even capture non-emotional features of the text in music. For example, recurring characters or locations in a novel could be indicated by recurring motifs. We will conduct human evaluations asking people to judge various aspects of the generated music such as the quality of music and the amount and type of emotion evoked by the music. We will also evaluate the impact of textual features such as the length of the novel and the style of writing on the generated music. Work on capturing note models (analogous to language models) from existing pieces of music and using them to improve the music generated by TransProse seems especially promising.

References

- [Agarwal et al.2011] Apoorv Agarwal, Boyi Xie, Iliia Vovsha, Owen Rambow, and Rebecca Passonneau. 2011. Sentiment analysis of twitter data. In *Proceedings of the Workshop on Languages in Social Media*, LSM '11, pages 30–38, Portland, Oregon.
- [Aisopos et al.2012] Fotis Aisopos, George Papadakis, Konstantinos Tserpes, and Theodora Varvarigou. 2012. Textual and contextual patterns for sentiment analysis over microblogs. In *Proceedings of the 21st International Conference on World Wide Web Companion*, WWW '12 Companion, pages 453–454, New York, NY, USA.
- [Ali and Peynirciolu2010] 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.
- [Alm and Sproat2005] Cecilia O. Alm and Richard Sproat, 2005. *Emotional sequencing and development in fairy tales*, pages 668–674. Springer.
- [Aman and Szpakowicz2007] Saima Aman and Stan Szpakowicz. 2007. Identifying expressions of emotion in text. In Vclav Matoušek and Pavel Mautner, editors, *Text, Speech and Dialogue*, volume 4629 of *Lecture Notes in Computer Science*, pages 196–205. Springer Berlin / Heidelberg.
- [Bakliwal et al.2012] Akshat Bakliwal, Piyush Arora, Senthil Madhappan, Nikhil Kapre, Mukesh Singh, and Vasudeva Varma. 2012. Mining sentiments from tweets. In *Proceedings of the 3rd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '12, pages 11–18, Jeju, Republic of Korea.
- [Balkwill and Thompson1999] 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.
- [Bellegarda2010] Jerome Bellegarda. 2010. Emotion analysis using latent affective folding and embedding. In *Proceedings of the NAACL-HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, Los Angeles, California.
- [Biles1994] John Biles. 1994. Genjam: A genetic algorithm for generating jazz solos. pages 131–137.
- [Bollen et al.2009] Johan Bollen, Alberto Pepe, and Huina Mao. 2009. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *CoRR*.
- [Boucouvalas2002] Anthony C. Boucouvalas. 2002. Real time text-to-emotion engine for expressive internet communication. *Emerging Communication: Studies on New Technologies and Practices in Communication*, 5:305–318.
- [Brody and Diakopoulos2011] Samuel Brody and Nicholas Diakopoulos. 2011. Cooooooooooooooooo!!!!!!!: using word lengthening to detect sentiment in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 562–570, Stroudsburg, PA, USA. Association for Computational Linguistics.
- [Brown1970] Calvin S Brown. 1970. The relations between music and literature as a field of study. *Comparative Literature*, 22(2):97–107.
- [Collins2008] Nick Collins. 2008. The analysis of generative music programs. *Organised Sound*, 13(3):237–248.
- [Cope1996] David Cope. 1996. *Experiments in musical intelligence*, volume 12. AR Editions Madison, WI.
- [Dowling and Harwood1986] W Jay Dowling and Dane L Harwood. 1986. *Music cognition*, volume 19986. Academic Press New York.
- [Gabrielsson and Lindström2001] Alf Gabrielsson and Erik Lindström. 2001. The influence of musical structure on emotional expression.
- [Genereux and Evans2006] Michel Genereux and Roger P. Evans. 2006. Distinguishing affective states in weblogs. In *Proceedings of the AAAI Spring Symposium on Computational Approaches to Analysing Weblogs*, pages 27–29, Stanford, California.
- [Hatten1991] Robert Hatten. 1991. On narrativity in music: expressive genres and levels of discourse in beethoven.
- [Hunter et al.2008] Patrick G Hunter, E Glenn Schellenberg, and Ulrich Schimmack. 2008. Mixed affective responses to music with conflicting cues. *Cognition & Emotion*, 22(2):327–352.
- [Hunter et al.2010] 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.
- [Iida et al.2000] Akemi Iida, Nick Campbell, Soichiro Iga, Fumito Higuchi, and Michiaki Yasumura. 2000. A speech synthesis system with emotion for assisting communication. In *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*.
- [Juslin and Laukka2003] Patrik N Juslin and Petri Laukka. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological bulletin*, 129(5):770.
- [Kim et al.2009] Elsa Kim, Sam Gilbert, Michael J. Edwards, and Erhardt Graeff. 2009. Detecting sadness in 140 characters: Sentiment analysis of mourning michael jackson on twitter.
- [Lerdahl and Jackendoff1983] Fred Lerdahl and Ray S Jackendoff. 1983. *A generative theory of tonal music*. MIT press.
- [Liu and Zhang2012] Bing Liu and Lei Zhang. 2012. A survey of opinion mining and sentiment analysis. In Charu C. Aggarwal and ChengXiang Zhai, editors, *Mining Text Data*, pages 415–463. Springer.

- [Liu et al.2003] Hugo Liu, Henry Lieberman, and Ted Selker. 2003. A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th International Conference on Intelligent User Interfaces*, pages 125–132, New York, NY. ACM.
- [Micznik2001] Vera Micznik. 2001. Music and narrative revisited: degrees of narrativity in beethoven and mahler. *Journal of the Royal Musical Association*, 126(2):193–249.
- [Mihalcea and Liu2006] Rada Mihalcea and Hugo Liu. 2006. A corpus-based approach to finding happiness. In *Proceedings of the AAAI Spring Symposium on Computational Approaches to Analysing Weblogs*, pages 139–144. AAAI Press.
- [Mohammad and Turney2010] Saif M. Mohammad and Peter D. Turney. 2010. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In *Proceedings of the NAACL-HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, LA, California.
- [Mohammad and Turney2013] Saif M. Mohammad and Peter D. Turney. 2013. Crowdsourcing a word-emotion association lexicon. 29(3):436–465.
- [Mohammad and Yang2011] Saif M. Mohammad and Tony (Wenda) Yang. 2011. Tracking sentiment in mail: How genders differ on emotional axes. In *Proceedings of the ACL Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*, WASSA '11, Portland, OR, USA.
- [Mohammad2012] Saif M. Mohammad. 2012. From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4):730–741.
- [Morrison and Demorest2009] Steven J Morrison and Steven M Demorest. 2009. Cultural constraints on music perception and cognition. *Progress in brain research*, 178:67–77.
- [Neviarouskaya et al.2009] Alena Neviarouskaya, Helmut Prendinger, and Mitsuru Ishizuka. 2009. Compositionality principle in recognition of fine-grained emotions from text. In *Proceedings of the Third International Conference on Weblogs and Social Media (ICWSM-09)*, pages 278–281, San Jose, California.
- [Pak and Paroubek2010] Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the 7th Conference on International Language Resources and Evaluation*, LREC '10, Valletta, Malta, May. European Language Resources Association (ELRA).
- [Pang and Lee2008] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in IR*, 2(1–2):1–135.
- [Perricone2000] J. Perricone. 2000. *Melody in Songwriting: Tools and Techniques for Writing Hit Songs*. Berklee guide. Berklee Press.
- [Picard2000] Rosalind W Picard. 2000. *Affective computing*. MIT press.
- [Pierre-Yves2003] Oudeyer Pierre-Yves. 2003. The production and recognition of emotions in speech: features and algorithms. *International Journal of Human-Computer Studies*, 59(1):157–183.
- [Roederer2008] Juan G Roederer. 2008. *The physics and psychophysics of music: an introduction*. Springer Publishing Company, Incorporated.
- [Rousseau2009] Jean-Jacques Rousseau. 2009. *Essay on the origin of languages and writings related to music*, volume 7. UPNE.
- [Schröder2009] Marc Schröder. 2009. Expressive speech synthesis: Past, present, and possible futures. In *Affective information processing*, pages 111–126. Springer.
- [Spencer1857] Herbert Spencer. 1857. The origin and function of music. *Frasers Magazine*, 56:396–408.
- [Strapparava and Valitutti2004] Carlo Strapparava and Alessandro Valitutti. 2004. WordNet-Affect: An Affective Extension of WordNet. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC-2004)*, pages 1083–1086, Lisbon, Portugal.
- [Tao and Tan2005] Jianhua Tao and Tieniu Tan. 2005. Affective computing: A review. In *Affective computing and intelligent interaction*, pages 981–995. Springer.
- [Tenney1988] James Tenney. 1988. *A history of consonance and dissonance*. Excelsior Music Publishing Company New York.
- [Thelwall et al.2011] Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2011. Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology*, 62(2):406–418.
- [Tumasjan et al.2010] Andranik Tumasjan, Timm O Sprenger, Philipp G Sandner, and Isabell M Welp. 2010. Predicting elections with twitter : What 140 characters reveal about political sentiment. *Word Journal Of The International Linguistic Association*, pages 178–185.
- [Turney and Littman2003] Peter Turney and Michael L Littman. 2003. Measuring praise and criticism: Inference of semantic orientation from association. *ACM Transactions on Information Systems*, 21(4).
- [Volkova et al.2010] Ekaterina P Volkova, Betty J Mohler, Detmar Meurers, Dale Gerdemann, and Heinrich H Bülhoff. 2010. Emotional perception of fairy tales: Achieving agreement in emotion annotation of text. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 98–106. Association for Computational Linguistics.
- [Webster and Weir2005] 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.
- [Zhe and Boucouvalas2002] Xu Zhe and A Boucouvalas, 2002. *Text-to-Emotion Engine for Real Time Internet Communication* *Text-to-Emotion Engine for Real Time Internet Communication*, pages 164–168.