Automatic Generation of Emotionally-Targeted Soundtracks

Kristine Monteith[1], **Virginia Francisco**[2], **Tony Martinez**[1], **Pablo Gervás**[2], **Dan Ventura**[1]
kristinemonteith@gmail.com, virginia@fdi.ucm.es, martinez@cs.byu.edu, pgervas@sip.ucm.es, ventura@cs.byu.edu

Computer Science Department[1]
Brigham Young University
Provo, UT 84602, USA

Departamento de Ingeniería del Software e Inteligencia Artificial[2]
Universidad Complutense de Madrid, Spain

Abstract

Music can be used both to direct and enhance the impact a story can have on its listeners. This work makes use of two creative systems to provide emotionally-targeted musical accompaniment for stories. One system assigns emotional labels to text, and the other generates original musical compositions with targeted emotional content. We use these two programs to generate music to accompany audio readings of fairy tales. Results show that music with targeted emotional content makes the stories significantly more enjoyable to listen to and increases listener perception of emotion in the text.

Introduction

Music has long been an integral aspect of storytelling in various forms of media. Research indicates that soundtracks can be very effective in increasing or manipulating the affective impact of a story. For example, Thayer and Levenson (l983) found that musical soundtracks added to a film about industrial safety could be used to both increase and decrease viewers' electrodermal responses depending on the type of music used. Bullerjahn and Guldenring (1994) similarly found that music could be used both to polarize the emotional response and impact plot interpretation. Marshall and Cohen (l988) noted significant differences in viewer interpretation of characters in a film depending on the type of accompanying music. Music can even affect the behavior of individuals after hearing a story. For example, Brownell (2002) found that, in several cases, a sung version of a story was more effective at reducing an undesirable target behavior than a read version of the story.

An interesting question, then, is whether computationally creative systems can be developed to autonomously produce effective accompaniment for various modalities. Dannenberg (1985) presents a system of automatic accompaniment designed to adapt to a live soloist. Lewis (2000) also details a "virtual improvising orchestra" that responds to a performer's musical choices. Similarly, our system is designed to respond to an outside entity when automatically generating music. Our efforts are directed towards providing accompaniment for text instead of a live performer.

This paper combines two creative systems to automatically generate emotionally targeted music to accompany the reading of fairy tales. Results show that emotionally targeted music makes stories significantly more enjoyable and causes them to have a greater emotional impact than music that is generated without regard to the emotions inherent in the text.

Methodology

In order to provide targeted accompaniment for a given story, each sentence in the text is first labeled with an emotion. For these experiments, selections are assigned labels of love, joy, surprise, anger, sadness, and fear, according the categories of emotions described by Parrot (2001). Selections can also be labeled as neutral if the system finds no emotions present. A more detailed description of the emotion-labeling system can be found in (Francisco and Hervás 2007). Music is then generated to match the labels assigned by the system. Further details on the process of generating music with targeted emotional content can be found in (Monteith, Martinez, and Ventura 2010).

Generating the actual audio files of the fairy tales with accompanying soundtrack was done following Algorithm 1. A text corpus is initially segmented at the sentence level (line 1) and each sentence is tagged with an emotion (line 2). Ten musical selections are generated for each possible emotional label and converted from MIDI to WAV format (lines 5-7) using WinAmp[1]. In order to produce a spoken version of a given fairy tale, each sentence is converted to an audio file (line 9) using FreeTTS,[2] an open-source text to speech program. This provides a collection from which musical accompaniments can be selected. Each audio phrase is analyzed to determine its length, and the musical file with matching emotional label that is closest in length to the sentence file is selected as accompaniment (lines 10-11). If all of the generated selections are longer than the audio file, the shortest selection is cut to match the length of the audio file. Since this is often the case, consecutive sentences with the same emotional label are joined before music is assigned

[1] http://www.winamp.com
[2] http://freetts.sourceforge.net

Algorithm 1 *Algorithm for automatically generating soundtracks for text.* F is the text corpus (e.g. a fairy tale) for which a soundtrack is to be generated.

SoundTrack(F)

1: Divide F into sentences: S_1 to S_m
2: Assign emotion labels to each sentence: L_1 to L_m
3: $S' \leftarrow$ join consecutive sentences in S with matching labels
4: $L' \leftarrow$ join consecutive matching labels in L
5: **for all** L'_i in L' **do**
6: Generate MIDI selections: M_{i1} to M_{i10}
7: Convert to WAV files: W_{i1} to W_{i10}
8: **for all** S'_i in S' **do**
9: $A_i \leftarrow$ Generate TTS audio recording from S'_i
10: $k \leftarrow argmin_j|len(A_i) - len(W_{ij})|$
11: $C_i \leftarrow A_i$ layered over W_{ik}
12: $O \leftarrow C_1 + C_2 + ... + C_n$
13: **return** O

(lines 3-4). Sentences labeled as "neutral" are left with no musical accompaniment. Finally, all the sentence audio files and their corresponding targeted accompaniments are concatenated to form a complete audio story (line 12).

Evaluation

Musical accompaniments were generated for each of the following stories: "The Lion and the Mouse," "The Ox and the Frog," "The Princess and the Pea," "The Tortoise and the Hare," and "The Wolf and the Goat." [3]

For comparison purposes, text-to-speech audio files were generated from the text of each story and left without musical accompaniment. (i.e. line 11 of Algorithm 1 becomes simply, $C_i \leftarrow A_i$.) Files were also generated in which each sentence was accompanied by music from a randomly selected emotional category, including the possibility of no emotion being selected (i.e. line 10 of Algorithm 1 becomes $k = rand(|L'| + 1)$, and file W_{i0} was silence for all i. Randomization was set such that $k = 0$ for approximately one out of three sentences.)

Twenty-four subjects were asked to listen to a version of each of the five stories. Subjects were divided into three groups, and versions of the stories were distributed such that each group listened to some stories with no music, some with randomly assigned music, and some with emotionally targeted music. Each version of a given story was played for eight people.

After each story, subjects were asked to respond to the following questions on a scale of 1 to 5: "How much did you enjoy listening to the story?" "If music was included, how effectively did the music match the events of the story?" and "Rate the intensity of the emotions (Love, Joy, Surprise, Anger, Sadness, and Fear) that were present in the story."

A Cronbach's alpha coefficient (Cronbach 1951) was calculated on the responses of subjects in each group to test for inter-rater reliability. Coefficients for the three groups were $\alpha = 0.93$, $\alpha = 0.87$, and $\alpha = 0.83$. (Values over 0.80 are generally considered indicative of a reasonable level of reliability and consequently, a sufficient number of subjects for testing purposes.)

[3] All audio files used in these experiments are available at http://axon.cs.byu.edu/emotiveMusicGeneration

	No Music	Random Music	Targeted Music
The Lion and the Mouse	2.88	2.13	2.75
The Ox and the Frog	3.50	2.75	3.00
The Princess and the Pea	3.00	3.38	4.13
The Tortoise and the Hare	2.75	2.75	3.88
The Wolf and the Goat	3.25	2.88	3.38
Average	3.08	2.78	3.43

Table 1: Average responses to the question "How much did you enjoy listening to the story?"

	Random Music	Targeted Music
The Lion and the Mouse	2.88	3.38
The Ox and the Frog	2.13	3.25
The Princess and the Pea	2.50	3.88
The Tortoise and the Hare	2.38	3.50
The Wolf and the Goat	1.75	3.25
Average	2.33	3.45

Table 2: Average responses to the question "How effectively did the music match the events of the story?"

Table 1 shows the average ratings for selections in each of the three categories in response to the question "How much did you enjoy listening to the story?" On average, targeted music made the selections significantly more enjoyable and random music made them less so. A Student's t-test reveals the significance level to be $p = 0.011$ for the difference in these two means. Selections in the "Targeted Music" group were also rated more enjoyable, on average, than selections in the "No Music" group, but the difference in means was not significant. Listeners did rate the version of "The Tortoise and the Hare" with emotionally targeted music as significantly more enjoyable than the "No Music" version ($p = 0.001$).

Table 2 reports the average ratings in response to the question "How effectively did the music match the events of the story?" Not surprisingly, music with targeted emotional content was rated significantly higher in terms of matching the events of the story than randomly generated music ($p = 0.003$).

Table 3 provides the intensity ratings for each of the six emotions considered, averaged over all five stories. Listeners tended to assign higher emotional ratings to selections in the "Random Music" category than they did to selections in the "No Music" category; however, this was not statistically significant. Average emotional ratings for the selections in the "Targeted Music" category had significantly higher ratings ($p = 0.027$) than selections accompanied by randomly generated music. When directly comparing "Targeted Mu-

	No Music	Random Music	Targeted Music
Love	1.83	1.40	1.55
Joy	2.03	2.10	2.53
Surprise	2.63	2.50	2.75
Anger	1.48	1.60	1.55
Sadness	1.60	1.70	2.05
Fear	1.58	2.00	2.15
Average	1.85	1.88	2.10

Table 3: Average intensity of a given emotion for all stories

	No Music	Random Music	Targeted Music
Love	1.75	1.38	1.75
Joy	2.03	2.10	2.53
Surprise	2.67	2.88	2.75
Anger	1.56	1.50	1.56
Sadness	1.94	2.06	2.31
Fear	1.94	2.13	2.31
Average	1.98	2.01	2.20

Table 4: Average intensity of labeled emotions for all stories

sic" with "No Music", average emotional ratings are again higher for the targeted music, though the difference falls a bit short of statistical significance ($p = 0.129$).

Table 4 gives average intensity ratings when only labeled emotions are considered (compare to Table 3). In this analysis, selections in the "Targeted Music" category received higher intensity ratings than selections in both the "No Music" and "Random Music" categories, with both differences being very near statistical significance ($p = 0.056$ and $p = 0.066$, respectively). Note that the only emotional category in which targeted music does not tie or exceed the other two accompaniment styles in terms of intensity ratings is that of "Surprise." The fact that "Random Music" selections were rated as more surprising than "Targeted Music" selections is not entirely unexpected.

Discussion and Future Work

Regardless of how creatively systems may behave on their own, Csikszentmihalyi (1996) argues that individual actions are insufficient to assign the label of "creative" in and of themselves. As he explains, "...creativity must, in the last analysis, be seen not as something happening within a person but in the relationships within a system." In other words, an individual has to interact with and have an impact on a community in order to be considered truly creative. Adding the ability to label emotions in text allows for generated music to be targeted to a specific project rather than simply existing in a vacuum.

In addition to allowing further interaction with the "society" of creative programs, our combination of systems also allows creative works to have a greater impact on humans. Music can have a significant effect on human perception of a story. However, as demonstrated in previous literature and in the results of our study, this impact is most pronounced when music is well-matched to story content. Music generated without regard to the emotional content of the story appears to be less effective both at eliciting emotion and at making a story more enjoyable for listeners.

Future work on this project will involve improving the quality of the generated audio files. Some of the files generated with the text-to-speech program were difficult to understand. A clearer reading, either by a different text-to-speech program or a recording of a human narrator, would likely enhance the intelligibility and possibly result in higher enjoyability ratings for the accompanied stories. Future work will also include adding more sophisticated transitions between musical selections in the accompaniment. This may also improve the quality of the final audio files.

Acknowledgments

This material is based upon work that is partially supported by the National Science Foundation under Grant No. IIS-0856089.

References

Brownell, M. D. 2002. Musically adapted social stories to modify behaviors in students with autism: four case studies. *Journal of Music Therapy* 39:117–144.

Bullerjahn, C., and Guldenring, M. 1994. An empirical investigation of effects of film music using qualitative content analysis. *Psychomusicology* 13:99–118.

Cronbach, L. J. 1951. Coefficient alpha and the internal structure of tests. *Psychometrika* 16(3):297–334.

Csikszentmihalyi, M. 1996. *Creativity: Flow and the Psychology of Discovery and Invention*. New York: Harper Perennial.

Dannenberg, R. 1985. An on-line algorithm for real-time accompaniment. *Proceedings of the Internation Computer Music Conference* 279–289.

Francisco, V., and Hervás, R. 2007. Emotag: Automated mark up of affective information in texts. In *EUROLAN 2007 Summer School Doctoral Consortium*, 512.

Lewis, G. 2000. Too many notes: Computers, complexity and culture in voyager. *Leonardo Music Journal* 10:33–39.

Marshall, S., and Cohen, A. J. l988. Effects of musical soundtracks on attitudes toward animated geometric figures. *Music Perception* 6:95–112.

Monteith, K.; Martinez, T.; and Ventura, D. 2010. Automatic generation of music for inducing emotive response. *Proceedings of the International Conference on Computational Creativity* 140–149.

Parrott, W. G. 2001. *Emotions in Social Psychology*. Philadelphia: Psychology Press.

Thayer, J., and Levenson, R. l983. Effects of music on psychophysiological responses to a stressful film. *Psychomusicology* 3:4454.