eMello: Auditory & Visual eBook Reader

Authors: TAN Zhao Tong, TAN Wei Jin, KHONG Jia Wei, Ivan CHRISTIAN, KOH Chuan Shen Marcus

Project Supervisor: ZHANG Kejun

Assistant Supervisors: CAO Yingping, DUAN Ke

ABSTRACT

This paper introduces the existing eBook and proposes possible improvements such as the addition of background music and dictionary overlays. A creation and implementation of such an eBook reading application for children is presented. We report the details of the design and creation of the application, followed by a discussion of potential improvements.

Keywords

Childhood learning; eBooks; Emotive learning; Music Emotion; Natural Language Processing; Music analysis

Introduction

Traditionally, children have been taught languages through the reading of books, even in their early stages of development. These books contain simple and easy to understand sentence structures and vocabulary. Books for younger children are also accompanied by simple illustrations that are easy on the eyes and have simple focal points for the child to develop their imagination and have a good imagery and understanding of the setting of the story. However, such a method only stimulates the visual senses and the child may not always get the full nuance that the writer is trying to portray.

To generate a deeper understanding of the narrative, an auditory accompaniment could be utilised. This taps on the success of film which utilises a visual heavy stimulation accompanied with an auditory stimulation to evoke emotion, build suspense and help the audience understand the narrative present. In a very similar manner, an eBook application provider, Booktrack, creates and publishes eBooks with the intention to make reading a more movie-like experience.

Objectives

Currently, Booktrack creates such a visualisation by providing music to create the atmosphere and sound effects to mimic environmental interactions. They also provide for a large target audience, yet still hold the same vision for all their books. In doing so, they are required to manually manage the input of the background music tracks and sound effect to their books.

However, in a page, there could be many different sound related words and thus many sound files would need to be loaded into the book. This assumes a person reads the auditory word when that particular sound effect is played. The feedback from Booktrack has shown that the overloading of sound effects is more of a distraction to users instead of an optimised tool.

We propose that a similar methodology can be used for the reading of children's books which focuses on the development of children below the age of 10. We targeted emotive learning through music instead of providing a movie like experience. By removing the sound effects, we aim to use the music to evoke emotive feelings in the child while increasing focus for reading yet still encourages imaginative development.

In addition, we provide the user a tool that aids in the development of their vocabulary. On top of which, we proposed the use of automated music sorting to provide a good set of tracks to be used per page.

Proposed Methodology

In order to achieve the above objective, we developed the following framework:

Using XCode, we built an iOS application that uses children's eBooks uploaded into it and by passing them through a Natural Language Processing (NLP) tool, IBM's Watson, we attained a tonal score.

Using the tone and score provided for each page, the application accesses the relevant music databases to attain a corresponding track to tag to each page of the eBook.

The final result is an eBook that plays a set track on each page. The book is then placed within the users' library for them enjoy and experience with the music.

Music Database

We were required to build up background instrumental music (BGM) databases that match the relevant tones contained by the Watson NLP and further divides these into sub-databases based on the degree of the tone (scores).

Through the python package for music and audio systems, libROSA, and using correlations to intensity, timbre, pitch and rhythm (Bhat et al, 2014), we were able to identify the mood of a song. The possible moods stem from Robert Thayer's model of mood, containing 8 different moods, presented in Figure 1.

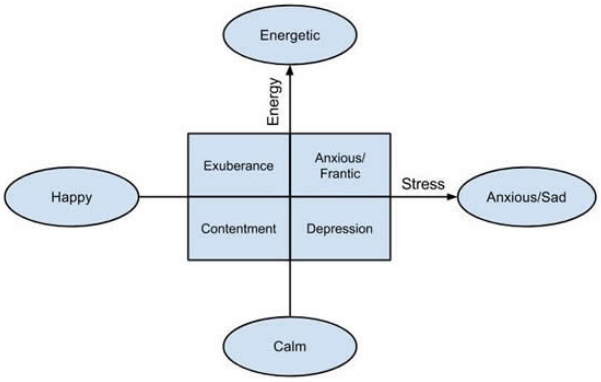

Figure 1. Robert Thayer's model of mood.

Using this python package, we trained and tested a music analyser model for us to analyse and sort music tracks into databases containing their classifications. We used the measurements of intensity, timbre, pitch and rhythm, all of which were derived from Bhat (2014) which are portrayed in Figure 2.

Mood	Intensity	Timbre	Pitch	Rhythm
Happy	Med	Med	V. High	V. High
Exuberant	High	Med	High	High
Energetic	V. High	Med	Med	High
Frantic	High	V. High	Low	V. High
Sad	Med	V. Low	V. Low	Low
Depression	Low	Low	Low	Low
Calm	V. Low	V. Low	Med	V. Low
Contentment	Low	Low	High	Low

Figure 2. Moods classified according to musical components. *Derived from Bhat (2014), obtained from Nuzzolo (2015).*

We understood that intensity is a measure of the loudness (in dB) of the piece and is a normalised value, while timbre is based off frequency responses over the note duration and is also normalised. On the other hand, pitch is based off the frequencies (in Hz) and rhythm is given in beats per minute (BPM) (Nuzzolo, 2015).

Based on the following classifications, we had trained our own mood analysing model to take the following into consideration and used data from both Bhat and Nuzzolo as references, comparing their data to our own. The data we obtained is shown in the following Figures 3 and 4. The figures contain the means values and standard deviation respectively.

Mood	Mean Intensity	Mean Timbre	Mean Pitch	Mean Rhythm
Happy	0.33070	0.30477	880.36	126.32
Exuberant	0.37077	0.34019	950.23	125.43
Energetic	0.40216	0.38048	1003.1	120.06
Frantic	0.31629	0.34938	1096.2	113.07
Sad	0.23189	0.21657	763.95	121.58
Depression	0.28555	0.21151	709.40	129.04
Calm	0.34021	0.29048	871.52	118.50
Contentment	0.30323	0.23233	807.10	121.66

Figure 3. Mean values attained from the Music Analyser.

Mood	Intensity	Timbre	Pitch	Rhythm
Happy	0.08892	0.08041	168.51	20.137
Exuberant	0.10215	0.10597	149.93	17.488
Energetic	0.09595	0.07266	170.96	12.903
Frantic	0.11111	0.09327	221.22	18.888
Sad	0.05253	0.06675	178.09	24.616
Depression	0.09809	0.06048	160.31	19.172
Calm	0.10745	0.13447	226.05	20.941
Contentment	0.08959	0.13447	214.29	20.336

Figure 4. Standard deviation values attained from the Music Analyser.

We then created a conversion table, Figure 5, to correlate the Watson Tone Analyser and the mood model. From the values attained via our music analyser and by calculating the confidence value, we classified the music pieces based on their moods.

Watson Tone Analyser	Robert Thayer
Anger	Energetic
Fear	Frantic
Joy	Happy + exuberant
Sadness	Anxious/sad
Analytical	Calm
Confident	Contentment
Tentative	Depression

Figure 5. eMello Tone-Mood conversion table.

By matching the confidence scores of both the tone and mood analysers, we were able to ensure a good match between the song and the page.

Book Libraries

When the app is released, users will be able to upload their own eBooks obtained from their respective sources and upload the .epub file into the application from the user's iCloud drive to our server. The uploaded file will then be converted into .html files.

These files will then be passed through the Watson Tone Analyser and this will tag each page to a tone and score. This score is a measure of the degree of the tone being presented on the page.

Through the relevant music libraries that we have created via the previous process mentioned, a song suiting the tone will be chosen as the accompanying track and tagged to that specific page. Once the process is completed, the book will be compiled back and available for the user to read from the Bookshelf page within the application.

Figure 6. Flowchart for the eMello Music Tagging Framework (MTF).

The same .html file will also run through our own word analyser, which isolates words that have two or more syllables.

These words are then run through a word corpus, the Corpus of Contemporary American English, and filters them through the top 5000 most common words used in the English Language. The words are also run through a list of basic vocabulary up to the 5th grade from Reading Rockets (Graham et al, 1993). Any word that is not within the 2 lists or are within the corpus and not in the Reading Rockets list but have more than 8 letters, will be identified as and classified under the difficult words.

This decision was made due to the commonality of the words being proportional to the probability the child would have come across the word previously. The less common or longer the word, the higher the probability the word will be foreign to the child.

After these difficult words are identified, a python dictionary containing all the difficult words contained within the book are compiled as keys. The meaning of these words are taken from the Oxford English Dictionary and added to the value of these keys.

A html script is then created to identify and write over the current .html files containing these words, creating popups that contain the explanation of the word. These popups will be available for the user to access if a said page contains the difficult word.

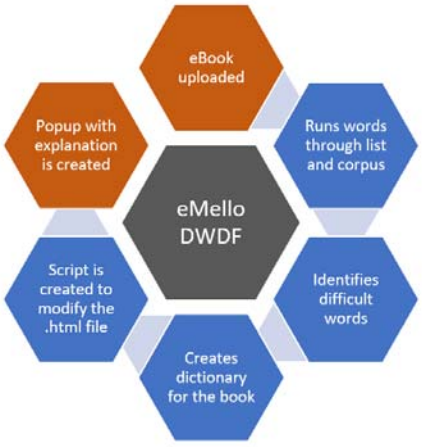

Figure 7. Flowchart for the eMello Difficult Word Dictionary Framework (DWDF).

Application Interface

The application user interface (UI) has been designed to be child-friendly, simple in design, with minimal input requirements by the user to get the desired output.

On the first opening of the application, a tutorial page will trigger and teach the user and their parent how to use the application.

On the start page there are 2 options: “Let’s Start Reading” and “Settings”.

Figure 8. eMello Start Page

In the settings page, users can toggle the “Default BGM” setting on and off. This will dictate whether the music will play by default when the book is opened. Additionally, this page includes credits for our application.

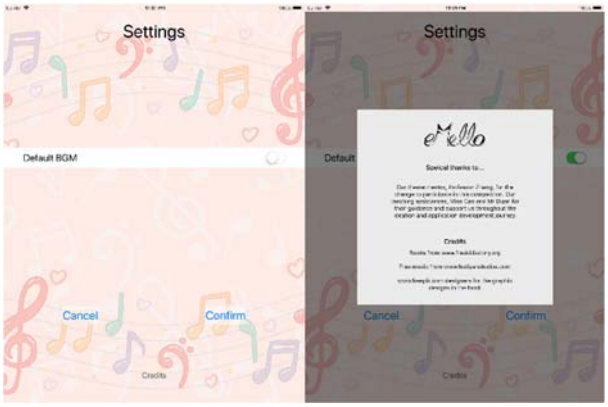

Figure 9. eMello settings page and credits

Pressing the “Let’s Start Reading” button will direct the user to the Bookshelf page, the users are presented with all the books for the user to read. For the books to appear on the library page, they first have to be uploaded via a “+” icon seen after all the available books. This will direct the user to his iCloud drive to obtain the .epub file.

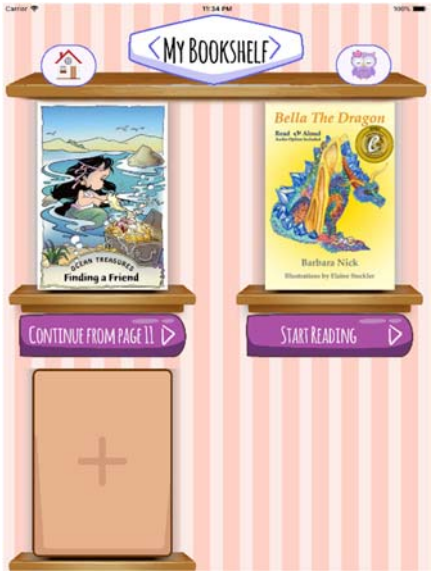

Figure 10. eMello My Bookshelf page

From the My Bookshelf page, clicking on a book cover will direct the user to its story. With the "Default BGM" setting on, the BGM associated with the first page will be played as the book is opened. Subsequently, the user can toggle the playing of the BGM on every page of the book.

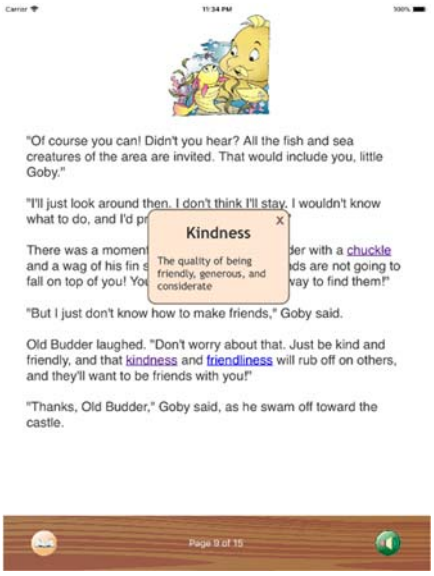

Figure 11. Sample page of the eBook within eMello

Within some of the pages of the book contain highlighted words. These highlighted words are difficult words as stated in the previous section. Upon seeing a highlighted word, the user may tap on it to generate a popup explaining the meaning of that word.

Finally, on each page within the book, there is a Bookshelf button at the bottom left hand corner which will automatically save the progress of the user while brining the user back to the Bookshelf page.

Discussion

We will discuss the intended course of action, potential benefits that pairing music to books can bring and possible improvements that could be implemented or changed in the future.

Considerations and Future Work

At this current stage, frameworks have been created and tested. The two frameworks work well in tandem but have not been fully automated due to the lack of a server platform to do so. The pseudo application created at this point is a proof of concept and the full automation of this system is still in development.

As a result, the application has yet to be released for beta testing and data collection with regards to the effectiveness of its intended purpose at this stage is not possible.

We thus plan to work with a server to run both frameworks in tandem and automate this process of creating the music integrated eBook. Ideally, the system will resemble that of what is presented in Figure 12 below.

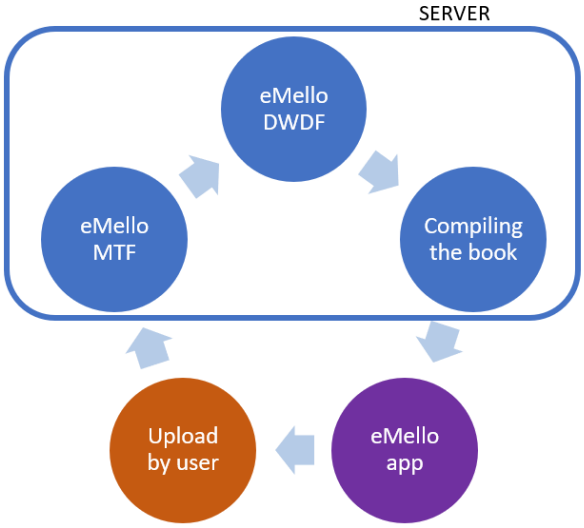

Figure 12. Macro view of the movement of the data of the eBook.

Music and Its Benefits

In the study, it was found that "top-down modulation of visual processing is not purely predictive in nature: mood, in this case manipulated by music, may also directly alter the way we perceive the world." (Jolij J., Meurs M., 2011) In a similar manner, we believe that in placing such mood provoking music can aid in the child's perception of the mood that the author wants to portray. This allows for a development of their emotive capacity alongside language development. This allows them to have a better emotive understanding of the characters within the story and the emotive environment the author intended from the scene.

As the BGM is an instrumental, it can also have the secondary effect to help the user concentrate and engage more with the text. Studies have shown that such background music can help in language development, and is echoed by researchers, who state that the mood created by the music "breaks down barriers and create an environment that is friendly." (Merrell A., 2004) This in turn makes it less of a daunting task to learn the language. Also, using the form of storybooks decreases the hostile-like environment and increases the level of understanding of difficult words within the prose.

Possible Improvements

Possible changes and further improvements in the design could include an upgrade of the dictionary and that of increasing the music continuity within the book.

A possible upgrade the dictionary component could get is by changing the purely text-based explanation of the popup dictionary component to one that contains pictures, sounds or short 5 second clips to demonstrate the meaning of the word. This will be beneficial since some words could be better explained by other means, as opposed to by text.

For example, sound words like "howl" could be explained as a word that means "to utter a loud, prolonged, mournful cry, as that of a dog or a wolf." However, the child may not be able to fully understand this description or imagine the sound. Attaching this sound clip to the popup will allow for them to have a better understanding and aid in a better visualisation and imagination of the stories they read.

For the music aspect, the improvement that could happen is that of the increasing music continuity. At the present moment, the music tracks are requested from Fesliyan Studios and Incompetech, and as a result the individual, non-continual tracks are attached to the individual pages. However, this system could be upgraded to eventually create an overarching loop track that runs through the whole book. With time, it would also be possible to have individual looped tracks on each page that harmonises well with this overarching track. This allows for a better continuity of the music which can be less disruptive when there is a change in the tone of the page and thus the songs. This would allow for a matching of the flow of the story without making it disruptive for the reader.

Acknowledgements

This work is supported by the International Design Institute in Zhejiang University and SUTD-ZJU Collaboration Office in Singapore University of Technology and Design. We would like to thank our theme mentor Prof. Zhang Kejun, Ms. Cao YingPing and Mr. Duan Ke for their guidance during this period. This project would not have been possible without their support and encouragement.

References

1. Bhat, A. S., V. S., A., S. Prasad, N., & Mohan D., M. (2014). *An efficient classification algorithm for music mood detection in western and Hindi music using audio feature extraction.* 2014 Fifth International Conference on Signal and Image Processing, pp. 359-364. Retrieved from

   http://dx.doi.org/10.1109/ICSP.2014.63

2. Corpus of Contemporary American English

   https://www.wordfrequency.info/free.asp

3. Fesliyan Studios

   www.fesliyanstudios.com

4. Graham S., Harris K. R., Loynachan C. (1993) *The Basic Spelling Vocabulary List.* Journal of Education Research 86(6) 363-368. Retrieved from

   http://www.readingrockets.org/article/basic-spelling-vocabulary-list

5. Hampiholi, V. (2012). *A method for Music Classification Based on Perceived Mood Detection for Indian Bollywood Music*. World Academy of Science, Engineering and Technology, International Science Index 72, International Journal of Computer, Electrical, Automation, Control and Information Engineering, 6(12), 1636 - 1643. Retrieved from

   https://waset.org/publication/15269

6. Jolij J., Meurs M. (2011). *Music Alters Visual Perception.* Retrieved from

   https://doi.org/10.1371/journal.pone.0018861

7. Kanters P. W. M. (2009). *Automated Mood Classification for Music.* Retrieved from

   http://arno.uvt.nl/show.cgi?fid=95615

8. Kevin MacLeod. Incompetech music.

   http://incompetech.com/

9. LibROSA Development Team (2013). *LibROSA python package.* Retrieved from

   http://librosa.github.io/librosa/

10. Merrell A. (2004). *The Benefits of Incorporating Music in the Classroom.* Retrieved from

    https://www.slideshare.net/vickytg123/the-benefits-of-incorporating-music

11. Nuzzolo M. (2015). *Music Mood Classification.* Retrieved from

    https://sites.tufts.edu/eeseniordesignhandbook/2015/music-mood-classification/

12. Oxford Dictionaries API

    https://developer.oxforddictionaries.com/

13. Tan S. L., Spackman M. P., Wakefield E. M. (2017). *The Effects of Diegetic and Nondiegetic Music on Viewer's Interpretations of a Film Scene.* Music Perception: An Interdisciplinary Journal, Vol. 34 No. 5, June 2017, pp. 605-623. Retrieved from

    http://dx.doi.org/10.1525/mp.2017.34.5.605/

14. Tan S. L. (2013). *How Film Music Shapes the Storyline.* Retrieved from

    https://www.psychologytoday.com/us/blog/what-shapes-film/201310/how-film-music-shapes-the-storyline

15. Rahman T. (2017). *Python Vocabulary 1.0.4* Retrieved from

    https://pypi.org/project/Vocabulary/1.0.4/

16. Watson Tone Analyser, IBM

    https://console.bluemix.net/docs/services/tone-analyzer/index.html#about