SentiBooks: Enhancing Audiobooks via Affective Computing and Smart Light Bulbs

Anna-Marie Ortloff
anna-marie.ortloff@student.ur.de
Media Informatics Group
University of Regensburg
Regensburg, Germany

Lydia Güntner
lydia-maria.guentner@student.ur.de
Media Informatics Group
University of Regensburg
Regensburg, Germany

Maximiliane Windl
maximiliane.windl@student.ur.de
Media Informatics Group
University of Regensburg
Regensburg, Germany

Thomas Schmidt
thomas.schmidt@ur.de
Media Informatics Group
University of Regensburg
Regensburg, Germany

Martin Kocur
martin.kocur@ur.de
Media Informatics Group
University of Regensburg
Regensburg, Germany

Christian Wolff
christian.wolff@ur.de
Media Informatics Group
University of Regensburg
Regensburg, Germany

ABSTRACT

We present *SentiBooks*, a smartphone application to enhance the audiobook listening experience via affective computing and smart light bulbs. Users can connect to *Philips Hue* Light Bulbs with a smartphone app while listening to an audiobook. The app analyzes the emotional expression of the narrator of the audiobook using speech emotion recognition and adjusts the colors of the lighting settings according to the expression of the narrator in 10-seconds intervals. By transitioning between colors that are connected to the specific emotion that is currently dominant in the reading, the overall audiobook experience is intensified.

CCS CONCEPTS

• **Human-centered computing → Visualization**; *Interaction paradigms*; *Ubiquitous and mobile computing*; • **Computing methodologies** → Machine learning.

KEYWORDS

affective computing, smart light bulbs, lighting, emotion detection, sentiment analysis

ACM Reference Format:
Anna-Marie Ortloff, Lydia Güntner, Maximiliane Windl, Thomas Schmidt, Martin Kocur, and Christian Wolff. 2019. SentiBooks: Enhancing Audiobooks via Affective Computing and Smart Light Bulbs. In *Mensch und Computer 2019 (MuC '19), September 8–11, 2019, Hamburg, Germany*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3340764.3345368

MuC '19, September 8–11, 2019, Hamburg, Germany

ACM ISBN 978-1-4503-7198-8/19/09.
https://doi.org/10.1145/3340764.3345368

1 INTRODUCTION

Audiobooks are a popular form of entertainment as recent numbers on the rapid revenue growth of downloadable audiobooks in the U.S. show [1]. Have and Pederson refer to the popularity of smartphones and online streaming services like *Spotify* and *Audible* to explain the growing expansion of the user group [6]. Indeed, the flexibility of these technologies led many users to listen to audiobooks in the car or while performing mundane tasks. Current surveys show that the vast majority of users listens to audiobooks at home [2]. In the following demonstration paper, we present our first explorations on how to enhance the audiobook listening experience in this specific home context. We present a novel approach to combine affective computing and the surrounding lighting system to create an illumination that is in line with the emotional expression of the current passage of the audiobook.

The illumination and lighting systems in our house are an important part of our everyday living. Research showed that illumination has an influence on our behavior and activities [18], our state of relaxation [10] and our attention [10]. With the rise of *smart home technology*, many everyday objects, including lighting systems and light bulbs also became "smart" and connected to the Internet. Users of smart light bulbs are able to change lighting settings and colors easily via smartphone apps. However, there are also applications allowing automatic adjustments of lighting depending on external factors such as the movement of the users [16] or the time of the day [12]. One of the few successful applications integrating colorful indoor lighting in everyday life are *Ambilight*

*Smart TV*s. *Ambilight Smart TV*s possess multiple smaller LED-bulbs at the rear part of the TV that glow in the color of the TV image depending on the position of the LED bulb on the TV. This type of TV enhancement indeed improves the overall viewing experience [3]. Jones et al. build upon this idea and explored different ways to use the surrounding lightning of the TV for projected illusions in the context of video games to improve the game experience [9]. While it is rather simple to refer to the colors of a TV image for enhancement, it is unclear how to decide upon the most fitting illumination colors in the context of audiobooks since there is no clear visual channel to derive the color from. We propose to use the emotional expression of the narrator of the audiobook to define the lighting composition and enhance the hearing experience since emotions and sentiments are substantial parts in the reception of narrative literature. Furthermore, by relying on research about which colors evoke specific emotions in humans [13, 20, 21], we hypothesize that by presenting colored light in line with the current emotion of the narration, similar emotions are invoked in the listener, thus intensifying their experience. There are a lot of exciting developments considering the usage of novel technologies like augmented reality or tangible interfaces to enhance the reading experience of traditional books [14]. In a similar way our approach employs novel technologies to improve the experience of audiobooks.

To extract emotions we employ methods from the application area of affective computing. Affective computing deals with the recognition and simulation of human emotions [19]. Due to advances in the context of machine learning and neural networks, state-of-the-art systems that predict human emotion and sentiments perform rather well with accuracies around 90% [19]. The most accurate and successful results can be found for the research areas of the prediction of sentiments and emotions in written text (also often called *Sentiment Analysis* [11]) and facial emotion recognition [7]. Nevertheless, there are also various systems for the prediction of emotion from human speech, implementing machine learning on corpora of human voice snippets and their emotional expression. There are various voice related features used for machine learning. Some examples are the pitch, pitch dynamics or the loudness. An overview about approaches and tools can be found in [5]. Although speech emotion analysis is currently outperformed by other modalities, speech-based systems still achieve accuracies of up to 80% [5] and have the advantage to be rather language independent. Since those systems are mostly evaluated with corpora representing everyday language and speakers, we assume that the general accuracy for our use case might be higher since narrators of audiobooks tend to put more emotion in their pronunciation for narrative and effect-specific purposes. At the same time, they typically are actors or professionally trained speakers for which speech-based recognition algorithms should work well.

Table 1: Default colors for emotions

Emotion	Lighting composition
Neutral	Neutral lighting (White)
Anger	Red lighting
Fear	Purple lighting
Sadness	Blue lighting
Happiness	Orange lighting

Considering the connection between the emotional expression and the colors of the lighting composition, there is no established research about which illumination colors are related with specific emotions and feelings. Therefore, we will refer to basic research from the context of color psychology [13, 21] and design recommendations [20] to decide upon which type of emotion is represented by which color by default. However, we also enable users to adjust emotion-color connections (see section 3).

2 THE SENTIBOOKS APPLICATION

The main parts of our system are (1) the *Philips Hue Light Bulbs*, (2) the *Vokaturi Speech Emotion Recognition Tool* and (3) an *Android* smartphone app. We name our application *SentiBooks* as a compound of the words sentiment and book.

Philips Hue Light Bulbs

The Philips Hue Light Bulbs[1] are a smart light bulbs system consisting of multiple bulbs and a bridge connecting the bulbs to WiFi. Via a system API, developers can create application and address the lights to change the saturation, brightness, overall color of the bulbs and call upon a limited range of effects like a “flash” or “color loop”. The colors of these light bulbs are based on the CIE color space and support a specific limited range of colors. Note that there are specific colors that cannot be represented by the bulbs (e.g. black) or that differ slightly from the general understanding of a color (e.g. green). The default setup of the Philips Hue system consist of three light bulbs. Based on color psychology [13, 21] and design recommendations (e.g.[20]) we defined five light compositions for the three lamps to represent specific emotions we analyze via speech emotion recognition. Table 1 summarizes the emotions and the dominant color for the respective lighting setting. Note, that these are only our proposed default settings. We also offer the functionality to change and adapt the color-emotion connection as the users sees fit via the smartphone app.

[1]https://www2.meethue.com/en-us/bulbs

Speech Emotion Recognition

To identify the emotional expression of the narrator we use *Vokaturi*. *Vokaturi*[2] is a tool for speech emotion recognition offering libraries for various platforms. We use the free version *OpenVokaturi*, which is considered the best free speech emotion recognition software on the market [5]. Given the audio of spoken language, *Vokaturi* calculates values for the four emotions anger, fear, sadness and happiness and for a neutral emotional value of the speech. Vokaturi uses two audio databases consisting of recordings of actors speaking sentences with different intended emotions [4, 8]. Vokaturi uses this databases as training data for machine learning with various acoustic features like the pitch, the intensity and the spectral slope. The emotional values Vokaturi produces range between 0 and 1 and sum up to one. We calculate the emotion of the audiobook in 10-seconds intervals by selecting the emotion that is reported with the highest value by Vokaturi for a specific interval.

Smartphone Application

We have developed a smartphone app for *Android* to perform the communication between the light bulbs and the emotion recognition as well as for the overall interaction with the lamps. After connecting with the *Philips Hue* Light Bulbs, users can start a "recording" similar to a default audio player. Users can choose to use an external audio stream, e.g. by starting an audiobook on the smartphone or on an audio player, or use a saved audiobook on the smartphone and listen to it via headphones. Every 10 seconds the app tries to identify the most dominant emotion of the narration and thus adjusts the color setting of the bulbs. The name of the most dominant emotion is also shown on the screen of the app (see figure 1).

Although emotions might change more frequent we chose 10-seconds intervals since shorter intervals proved too fast and hectic while 10-seconds intervals lead to rather smooth transitions. However, in future evaluations we want to analyze which intervals work best for the user. Users can also save the bulb instructions of the recording and visualize the color progression of an audiobook thus allowing the analysis and comparison of the emotional development in audiobooks (see figure 2).

As mentioned, the research about color-emotion connections is not fully clear and dependent of the individual interpretation of the user [13]. Therefore, via the settings options of the smartphone app users can adjust the color-emotion connection, as they prefer (see figure 3).

[2]https://developers.vokaturi.com/getting-started/overview

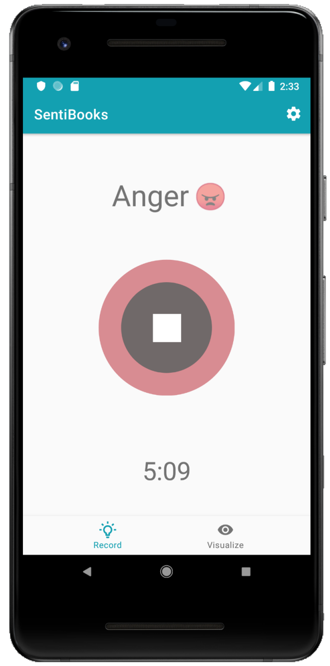

Figure 1: Emotion recognition while playing an audiobook

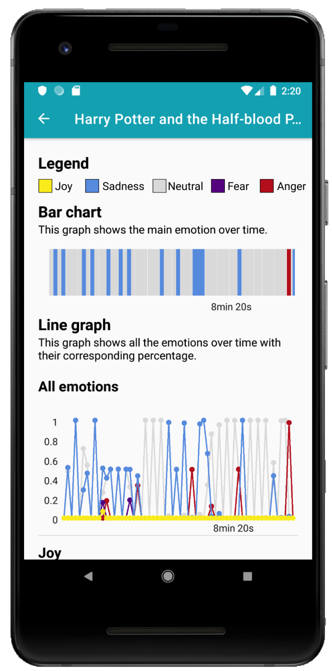

Figure 2: Emotion/color progression of a recording

3 DISCUSSION

The presented application and procedure is currently in an early stage of development. We just started first evaluations with 6 persons (3 female, 3 male with an average age of

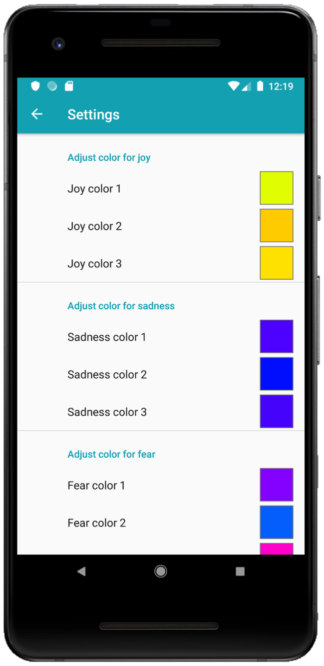

Figure 3: Changing the default color settings

40 years) that are regular audiobook listener. For the first evaluations we focused on usability issues of the app but we also gathered qualitative data about the overall impression via "thinking aloud". The feedback of the participants is mixed in the sense that users can imagine using the app but criticize that the emotion recognition seems arbitrary at times. Indeed, according to *Vokaturi* the accuracy for the free version of the emotion recognition is around 66%, which is rather low for real life applications. We plan to perform more evaluations with larger sample sizes focusing on the overall acceptance and experience of our implementation.

Furthermore, we also plan to improve the emotion recognition (1) by exploring the performance of alternative software packages [5] and (2) by combining the current approach with speech-to-text and textual sentiment analysis thus performing multimodal emotion analysis. Multimodal emotion analysis have been proven to outperform unimodal approaches for similar use cases [15] and has also found its way into the analysis of narrative texts [17]. Additionally, we currently solely focus on the color of the bulbs neglecting other aspects like the brightness or effects like flashing and transitions. We plan to integrate more aspects of the lighting system but also user-specific aspects (like the position of the user) to further enhance the listening experience.

REFERENCES

[1] AAP. 2018. AAP Reports Trade Revenues Up 6.3%, eBook Revenues Down 3.2% in First Quarter. https://the-digital-reader.com/2018/05/28/aap-reports-trade-revenues-up-6-3-ebook-revenues-down-3-2-in-first-quarter/
[2] Audio Publishers Association. 2018. Another banner year of robust growth for the audiobook industry. https://www.audiopub.org/uploads/pdf/2018-Consumer-Sales-Survey-Final-PR.pdf
[3] SHA Begemann. 2005. A scientific study on the effects of Ambilight in flat-panel displays. https://iristech.co/wp-content/uploads/2016/10/ambilight.pdf
[4] Felix Burkhardt, Astrid Paeschke, Miriam Rolfes, Walter F Sendlmeier, and Benjamin Weiss. 2005. A database of German emotional speech. In *Ninth European Conference on Speech Communication and Technology*.
[5] Jose Garcia-Garcia, Victor Penichet, and María Lozano. 2017. Emotion detection: a technology review. In *Proceedings of the XVIII International Conference on Human Computer Interaction*. 1–8. https://doi.org/10.1145/3123818.3123852
[6] Iben Have and Birgitte Stougaard Pedersen. 2015. *Digital audiobooks: New media, users, and experiences*. Routledge.
[7] Eva Hudlicka. 2003. To feel or not to feel: The role of affect in human–computer interaction. *International journal of human-computer studies* 59, 1-2 (2003), 1–32.
[8] P Jackson and S Haq. 2014. Surrey audio-visual expressed emotion (savee) database. *University of Surrey: Guildford, UK* (2014).
[9] Brett R Jones, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. 2013. IllumiRoom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 869–878.
[10] ES Lee and HJ Suk. 2012. The emotional response to lighting hue focusing on relaxation and attention. *Journal of Korean Society of Design Science* 25, 2 (2012), 27–39.
[11] Bing Liu. 2016. *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
[12] Remco Magielse and Serge Offermans. 2013. Future lighting systems. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. ACM, 2853–2854.
[13] KAYA Naz and H Helen. 2004. Color-emotion associations: Past experience and personal preference. In *AIC 2004 Color and Paints, Interim Meeting of the International Color Association, Proceedings*, Vol. 5. Jose Luis Caivano, 31.
[14] Mohammad Obaid, Ilgim Veryeri Alaca, Pawel W Wozniak, Lars Lischke, and Mark Billinghurst. 2017. Transforming Books and the Reading Experience through Interactive Technologies: Preface. *IxD&A: Interaction Design and Architecture (s)* 32 (2017), 71–73.
[15] Soujanya Poria, Erik Cambria, Rajiv Bajpai, and Amir Hussain. 2017. A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion* 37 (2017), 98–125.
[16] Esben Skouboe Poulsen, Hans Jørgen Andersen, Ole B Jensen, Rikke Gade, Tobias Thyrrestrup, and Thomas B Moeslund. 2012. Controlling urban lighting by human motion patterns results from a full scale experiment. In *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 339–348.
[17] Thomas Schmidt, Manuel Burghardt, and Christian Wolff. 2019. Toward Multimodal Sentiment Analysis of Historic Plays: A Case Study with Text and Audio for Lessing's Emilia Galotti.. In *DHN*. 405–414.
[18] Hyeon-Jeong Suk, Geun-Ly Park, and Yoon-sook Kim. 2012. Bon Appetit! An Investigation About Color Combination Between Lighting and Food. *Journal of Literature and Art Studies* 2, 5 (2012), 559–566.
[19] Jianhua Tao and Tieniu Tan. 2005. Affective computing: A review. In *International Conference on Affective computing and intelligent interaction*. Springer, 981–995.
[20] Frank Thissen. 2000. *Screen-Design-Handbuch: effektiv informieren und kommunizieren mit Multimedia*. Springer.
[21] Patricia Valdez and Albert Mehrabian. 1994. Effects of color on emotions. *Journal of experimental psychology: General* 123, 4 (1994), 394.