EMOLight: Immersive Visual Experience for the Audibly Impaired

AL TAMIMI, Abdel-Karim and BROCK, Matthew

EMOLight: Immersive Visual Experience for the Audibly Impaired

Abdel-Karim Al-Tamimi
School of Computing and Digital Technologies
Sheffield Hallam University
Sheffield, UK
a.al-tamimi@shu.ac.uk

Matthew Brock
School of Computing and Digital Technologies
Sheffield Hallam University
Sheffield, UK
b9013696@student.shu.ac.uk

Abstract— **In this paper, we introduce EMOLight, an innovative AI-driven ambient lighting solution that enhances viewer immersion, especially for the audibly impaired, by dynamically synchronising with the emotional content of audio cues and sounds. Our proposed solution leverages YAMNet deep learning model and Plutchik's emotion-colour theory to provide real-time audio emotion recognition, user-specific customisation, and multi-label classification for a personalised and engaging experience. This synchronisation enriches the viewing experience by making it more engaging and inclusive. This research demonstrates the feasibility and potential of EMOLight, paving the way for a future where technology adapts to diverse sensory needs and preferences, revolutionising the way we experience and interact with media.**

Keywords—Ambient Lighting, Affective Computing, Audio Emotion Recognition, Machine Learning, Multisensory Enhancement, Human-Computer Interaction

I. INTRODUCTION

Television manufacturers have increasingly incorporated ambient lighting technology to enhance viewer immersion, primarily by matching the ambient light to the colour profile displayed on the screen. This technique effectively extends the visual content beyond the physical bounds of the screen, creating a more enveloping viewing experience. Studies have demonstrated that such enhancements can significantly impact viewer engagement and emotional response [1], [2]. Building on this foundation, we propose in this paper an AI-driven solution that extends the concept of ambient lighting to include auditory elements. Our proposed solution, EMOLight, utilises pre-trained machine learning models to analyse and classify audio cues and sound bites extracted from soundtracks and scores in real time. By integrating Robert Plutchik's emotion-colour theory [3], shown in Figure 1, our solution dynamically enables the adjustment of the ambient lighting to reflect the emotional tone conveyed by the detected audio, thereby enriching the emotional resonance of the viewing experience.

The Plutchik emotion wheel, a seminal model conceived by psychologist Robert Plutchik, serves as a foundational framework for comprehending and visually representing the intricate tapestry of human emotions [4]. This influential conceptualization organizes emotions into primary, secondary, and tertiary categories, elegantly capturing the interrelationships and polarities that exist within the emotional spectrum. The wheel's principles of proximity and opposition have rendered it an invaluable tool across myriad domains. Its applications span sentiment analysis of vast social media datasets [5], enabling insights into the collective emotional landscape. Moreover, it has facilitated the development of user-friendly emotional tagging interfaces [6], empowering technology users to intuitively express and communicate their affective states. Researchers have also leveraged the wheel's structure to enhance emotion distribution learning through innovative label enhancement methodologies [7], refining the granularity and accuracy of emotional classification [8]. Researchers have also emphasised the importance of music in expressing and managing emotions, highlighting its potential for therapeutic and educational applications. Furthermore, concerted efforts have been undertaken to automatically annotate affective lexicons with nuanced intensity values using the WordNet Similarity software package [9].

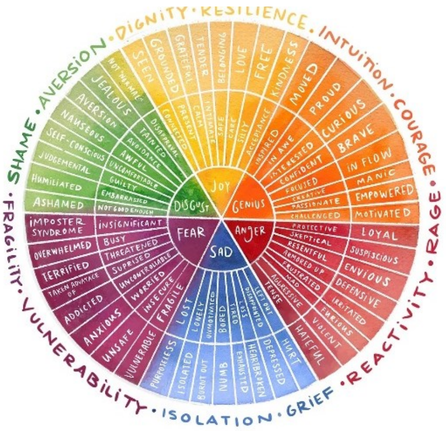


Fig. 1. The Plutchik's Emotion Wheel [10]

The relationship between colours and emotions is a complex topic that spans multiple disciplines, including psychology, marketing, and art. Associations between colours and emotions are influenced by factors such as hue, lightness, and saturation, with interpretations varying based on cultural and individual differences. Notably, chroma, or the intensity of colour, plays a crucial role in emotional perception. Images with higher chroma levels are typically associated with feelings of happiness and arousal, while lower chroma levels are more likely to evoke calmness or sadness [11].

Furthermore, a compelling study by [12] demonstrated that both colour-blind and non-colour-blind men attributed similar emotions to colours, regardless of whether those colours were presented through verbal descriptions or visual patches. Importantly, the intensity of the emotions reported was not influenced by the severity of the participants' colour blindness.

Our approach importantly acknowledges that emotional responses to colour are highly subjective, varying not only between individuals but also across different cultures and environments [13]. To account for this variability, we incorporated user-specific adjustments, allowing the emotion-colour associations to be customized based on individual preferences and personal perceptions. This customization is particularly aimed at enhancing accessibility for the audibly impaired, providing them with an alternative sensory channel

979-8-3315-1645-1/24/$31.00 ©2024 Crown

to perceive and engage with the emotional undertones of audiovisual content.

Our project aims to advance immersive visual and TV technologies by integrating audio-induced emotional cues into ambient lighting, creating a more inclusive and emotionally engaging viewing experience. This contributes to the growing field of multisensory enhancement technologies, which are vital for designing accessible, user-centric media systems.

II. Research Methodology

Our methodology involved a systematic approach, combining the review of existing machine learning models and relevant literature with the implementation of advanced audio processing techniques to develop our solution. Our findings focused on YAMNet [14] recognising its proven excellence in audio feature classification. Notably, the YAMNet deep learning model has showcased its adaptability and efficacy across a diverse range of applications, making it an attractive candidate for further exploration. Researchers have leveraged YAMNet to develop an acoustic detection system (ADS) for identifying emergency events, such as people screaming, explosions, or gunshots, in search and rescue operations. This ADS system achieved accuracies of 70.13% indoors and 70.52% outdoors [15]. Beyond emergency response, YAMNet has also been adapted for estimating human eating behaviours in real-time achieving high accuracy levels of 93.3% in frame-level classifications of eating behaviours, which can aid in monitoring and promoting healthier eating habits [16]. Researchers also utilized YAMNet to detect water-related activities in smart homes, such as flushing and showering, by analysing acoustic data collected from bathrooms. This approach was instrumental in developing a system capable of monitoring the daily routines of elderly individuals, thus enhancing home safety and healthcare efforts. The system demonstrated a classification accuracy of over 80% for water sounds using transfer learning methods on a Raspberry Pi-based edge device [17].

Concurrently, we reviewed academic literature on music genre classification, analysing studies that utilized both well-known public datasets like GTZAN [18] and specialized private collections [19]. This dual approach provided a comprehensive understanding of the state-of-the-art in audio classification and the contemporary best practices in audio processing. As a result, we adopted several pre-processing techniques to refine our solution’s audio input for better model performance. These included slicing audio samples into manageable segments, applying various audio filters to reduce noise and enhance signal quality, adjusting the sampling rate to optimal levels for digital processing, and converting stereo audio tracks to mono to standardize the input data. This allowed us to capture the nuances of chords, rhythm, melody, and dynamic patterns in the audio data.

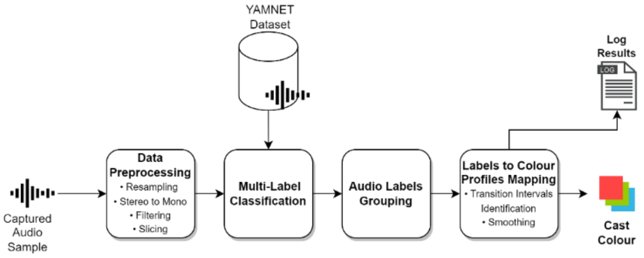


Fig. 2. Overview of EMOLight Processing Pipeline

The duration of audio slices was meticulously optimized to balance between temporal resolution and computational efficiency, ensuring the clarity and reliability of feature identification. Leveraging the pre-trained YAMNet model, our system classifies the audio cues and assign them multi-labels based on the audio features detected. A key challenge we faced in implementing our solution was managing the complexity of emotional responses to music, as outlined by Plutchik's emotion wheel. Our solution had to account for the inherent ambiguity of musical emotions, where a single sound could evokes multiple, overlapping emotional states, requiring a nuanced approach to capture this emotional diversity accurately.

To address this challenge, we devised a novel approach that effectively translates audio labels detected by YAMNet into corresponding emotional colours. This process leverages the diversity of human opinions to introduce variance, acknowledging the inherent subjectivity in interpreting audio cues as colour responses. Additionally, we employed OpenAI's ChatGPT to minimize bias and detect nuanced variations in word choices, which is crucial for accurately differentiating between closely related emotional cues. Future research could explore the use of other large language models (LLMs) to compare their performance and foster a deeper understanding of how LLMs perceive emotions, ultimately leading to a more robust consensus among LLMs.

To further enhance the emotional interpretation, we integrated the NRC Word-Emotion Association Lexicon (EmoLex) [20], a widely-used and well-established emotional lexicon that provides a comprehensive framework for understanding emotional associations. EmoLex has been extensively validated and has demonstrated high accuracy in capturing emotional nuances, with studies showing that it can effectively distinguish between subtle emotional differences By leveraging EmoLex, we were able to tap into its vast repository of emotional knowledge, which is based on a large-scale crowdsourcing effort involving over 2,000 participants [21]. This allowed us to create a more robust and accurate emotional interpretation system, which was further enhanced by our mixture-of-experts approach that synthesized insights from human experts, ChatGPT and EmoLex.

Table I shows a sample of YAMNet audio labels and the defined consensus labels. As shown, the EMOLex lexicon doesn't always align with YAMNet's labels and may assign multiple emotions to a single label, highlighting the complexity of emotional interpretation. Future work will explore multiple emotion associations to better capture emotional expression and deepen our understanding of the connections between emotions and labels.

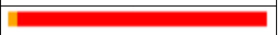
TABLE I. Audio-Label to Emotion Translation

YAMNet Label	*Human Expert 1*	*Human Expert 2*	*ChatGPT*	*EmoLex*	***Consensus***
Yell	Anger	Fear	Anger	Anger / Fear	Anger
Bellow	Anger	Anger	Anger	Anger	Anger
Laughter	Joy	Joy	Joy	Joy	Joy
Giggle	Joy	Joy	Joy	Joy	Joy
Crying	Sadness	Sadness	Sadness	Sadness	Sadness
Baby Crying	Sadness	Sadness	Sadness	-	Sadness
Mantra	Fear	Fear	Joy	-	Fear
Groan	Disgust	Disgust	Sadness	Disgust	Disgust

III. IMPLEMENTATION AND PRELIMINARY RESULTS

To evaluate the program's performance in identifying emotions, we conducted several tests using 10 royalty-free songs from Pixabay [17], spanning various genres and artists. Each song was analysed to assess the emotions it evoked, and a *main-colour* representing the dominant emotion was assigned for quick visual comparison. This method aimed to provide a reproducible and qualitative assessment of the program's accuracy in matching perceived subjective emotions with corresponding colours. By integrating these emotional cues with colour representation, we created an intuitive way to visualize the emotional landscape of each song. The results are displayed in Table II below, highlighting the program's effectiveness in aligning musical emotions with their visual counterparts.

TABLE II. EMOLIGHT TESTING RESULTS – PIXABAY SAMPLES

Song Name	*Expected Main Colour(s)*	*EMOLight Results*
Smoke.wav	Orange	
Risk.wav	Orange/Red	
Price_of_Freedom.wav	Orange/Blue	
Dark_matter.wav	Red	
Piano_Moment.wav	Blue/Orange	
Mountain_Path.wav	Orange/Yellow	
Futuristic_Beat.wav	Yellow/Orange	

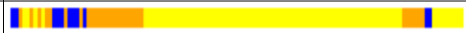
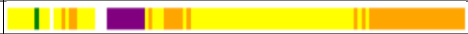
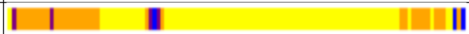
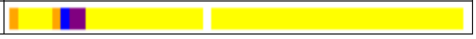
We further evaluated EMOLight's performance by analysing the 1967 film *The Jungle Book*, chosen for its diverse musical scores, emotional depth, and variety of character-driven scenes, offering an expansive dataset to rigorously assess EMOLight's capabilities. The film was divided into eleven segments, each approximately ten minutes long, to enhance data clarity and facilitate in-depth interpretation. This segmentation enabled a comprehensive analysis of emotional transitions throughout the film, allowing us to capture the subtle emotional nuances embedded within the narrative and inter-character interactions.

The comprehensive analysis presented in Table III highlights the nuanced portrayal of emotions within the cinematic narrative, as illustrated by the use of color-coded visualizations. Joyful moments, such as Mowgli's light-hearted interactions with Baloo the bear, are predominantly represented by vibrant yellow tones that exude warmth and happiness. These scenes evoke a sense of pure delight and amusement, capturing the essence of youthful camaraderie.

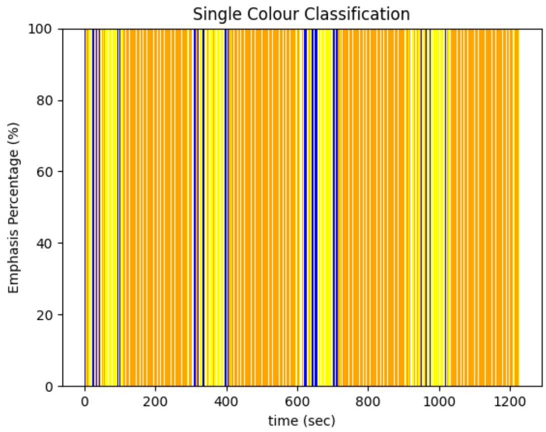
Conversely, moments of peril or tension, particularly during encounters between Mowgli and aggressive predators such as Shere Khan, are characterized by deep purple hues. This colour scheme effectively conveys a spectrum of emotions ranging from fear to suspense, highlighting the intensity and urgency of these dramatic events. Sad and melancholic scenes, notably the poignant separation of Mowgli from his adoptive wolf family, are depicted in a blue palette. These tones encapsulate feelings of sorrow, loss, and longing, resonating with the emotional depth of such pivotal moments within the film's storyline.

Additionally, white areas on the color-coded timeline indicate periods where there is no accompanying background music or sound effects, solely featuring speech dialogues. This suggests gaps in the narrative that may benefit from further integration into EMOLight that could better interpret and capture the subtle nuances of spoken language to enrich its overall capabilities.

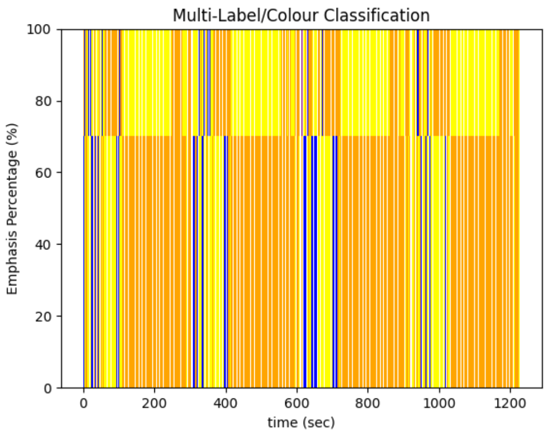
TABLE III. EMOLIGHT TESTING RESULTS – THE JUNGLE BOOK 1967

Film Segment	*EMOLight Results*
Jungle_Book_1	
Jungle_Book_2	
Jungle_Book_3	
Jungle_Book_4	
Jungle_Book_5	
Jungle_Book_6	
Jungle_Book_7	
Jungle_Book_8	
Jungle_Book_9	
Jungle_Book_10	
Jungle_Book_11	

This example demonstrates the effectiveness of EMOLight in capturing and visualizing the emotional landscape of a film. By translating audio cues into corresponding colours, EMOLight provides a valuable tool for understanding and analysing the emotional impact of audio in film and other media.



(a) Single colour/emotion identification for *Price of Freedom* Audio Sample



(b) Multi colour/emotion identification for *Price of Freedom* Audio Sample

Fig. 3. Single Label vs. Multi-Label/Colour Classification

Based on YAMNet capability of detecting multiple sounds and label them accordingly, we further explored the application of multi-label classification model by highlighting multiple colours based on the dominant emotions identified.

The results, as depicted in Figure 3, illustrate how multiple emotions can be detected in a test sample, audio sample labelled *Price of Freedom*, indicating a potential mixture of emotions, with the most dominant colour/emotion receiving greater emphasis (in this case we represented the top dominant emotion by 70% of the stacked bar, and 30% for the second identified emotion).

To enhance the representation of these mixed emotions or colours, we are currently exploring several approaches. One approach is to implement a gradient blending technique that visually represents the degree of each emotion by varying the intensity and saturation of the corresponding colours. Another method involves using colour overlays with varying opacities that reflect the proportionate presence of each detected emotion within the sample. Additionally, interactive visual elements could be introduced, allowing users to adjust the visibility of each emotion to explore their interplay more thoroughly. These methods would not only provide a more nuanced representation of emotional complexity but also enhance interpretability for end users.

IV. Conclusions

This paper introduced our AI-driven ambient lighting solution, viz. EMOLight, that synchronizes colours with audio content to enhance viewer immersion and accessibility based on emotional cues. EMOLight was tested on a variety of samples showcasing the diverse emotional spectrum in these samples and especially within the 1967 film, *The Jungle Book*. Results showed that EMOLight could effectively capture and visualize emotions across different media types. We also showcased the potentials of EMOLight as a multi-label classification model which allows highlighting different colours based on dominant emotions identified. This approach helps users perceive complex emotional states by indicating the presence of more than one emotion simultaneously.

This approach with applying the appropriate colour blending techniques and interactive visual elements present tremendous potential enhancements to represent mixed emotions, aiming for a more nuanced depiction that can adjust user preferences.

Conclusively, EMOLight demonstrates the feasibility of integrating technology into media experiences to adapt sensory needs based on emotional content. This research opens new avenues for further development in sensory interaction technologies, potentially setting standards and leading future advancements in accessible media experiences. Future work includes expanding emotion recognition capabilities, conducting user studies, exploring integration with other media formats, enhancing visualization techniques, and investigating how EMOLight can be applied to a wider range of media types.

Acknowledgement

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version of this paper arising from this submission.

References

[1] A. Löcken, A.-K. Frison, V. Fahn, D. Kreppold, M. Götz, and A. Riener, "Increasing user experience and trust in automated vehicles via an ambient light display," in Proc. 22nd Int. Conf. Human-Computer Interaction with Mobile Devices and Services, 2020, pp. 1-10.

[2] A. Weffers-Albu, S. De Waele, W. Hoogenstraaten, and C. Kwisthout. "Immersive TV viewing with advanced Ambilight." In 2011 IEEE International Conference on Consumer Electronics (ICCE), pp. 753-754. IEEE, 2011.

[3] R. Plutchik, "The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," American Scientist, vol. 89, no. 4, pp. 344-350, 2001.

[4] A. Semeraro, S. Vilella, and G. Ruffo, "PyPlutchik: Visualising and comparing emotion-annotated corpora," PLOS ONE, 2021. [Online]. Available: https://doi.org/10.1371/journal.pone.0256503

[5] A. Mondal and S. S. Gokhale. "Mining emotions on Plutchik's wheel." In 2020 seventh international conference on social networks analysis, management and security (snams), pp. 1-6. IEEE, 2020.

[6] K. Warpechowski, D. Orzeszek, and R. Nielek, "Tagging emotions using a wheel user interface," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland Uk, May 2019, pp. 1-12, doi: 10.1145/3351995.3352056.

[7] X. Zeng, Q. Chen, S. Chen, and J. Zuo, "Emotion Label Enhancement via Emotion Wheel and Lexicon," Mathematical Problems in Engineering, vol. 2021, pp. 1-12, 2021, doi: 10.1155/2021/6695913.

[8] R. S. Hirschprung, O. Leshman, and C. Hajaj. "The Sound of Emotions: An Artificial Intelligence Approach to Predicting Emotions from Musical Selections." , Dec 2023, Available at SSRN 4656595, doi: dx.doi.org/10.2139/ssrn.4656595.

[9] C. Molina-Beltrán et al., "Improving the affective analysis in texts: Automatic method to detect affective intensity in lexicons based on Plutchik's wheel of emotions," The Electronic Library, vol. 37, no. 6, pp. 984-1006, 2019, doi: 10.1108/EL-11-2018-0219.

[10] A. van Muijen, "Watercolor Emotion Wheel," AvanMuijen.com, 2024. [Online]. Available: https://www.avanmuijen.com/watercolor-emotion-wheel. [Accessed: Jul. 13, 2024].

[11] A. D. Pazda, et al. "Colorfulness influences perceptions of valence and arousal." Journal of Experimental Psychology: General 153, no. 1 (2024): 145.

[12] D. Jonauskaite et al. "Colour-emotion associations in individuals with red-green colour blindness PeerJ. 2021 Apr 7;9:e11180. doi: 10.7717/peerj.11180. PMID: 33868822; PMCID: PMC8035895.

[13] J. Lyu, "Analysis of the Effect of Different Types of Colors on Human Behavior and Emotion," in 2022 3rd International Conference on Mental Health, Education and Human Development (MHEHD 2022), pp. 1060-1063, Atlantis Press, 2022.

[14] M. Plakal and D. Ellis, "YAMNet," July 2024. [Online]. Available: https://www.tensorflow.org/hub/tutorials/yamnet

[15] T. Marinopoulou, A. Lalas, K. Votis, and D. Tzovaras, "An AI-powered Acoustic Detection System Based on YAMNet for UAVs in Search and Rescue Operations," in NOISE-CON ... proceedings, 2023, doi: 10.3397/in_2023_0749.

[16] W. Chen, H. Kamachi, A. Yokokubo, and G. Lopez, "Bone Conduction Eating Activity Detection based on YAMNet Transfer Learning and LSTM Networks," pp. 74-84, doi: 10.5220/0010903700003123, 2022.

[17] S. H. Hyun, "Sound-Event Detection of Water-Usage Activities Using Transfer Learning." Sensors 24, no. 1 (2023): 22.

[18] B. L. Sturm, "The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use," arXiv preprint arXiv:1306.1461, 2013.

[19] A. Al-Tamimi, M. Salem, and A. Al-Alami, "On the use of feature selection for music genre classification," in Proc. 2020 Seventh Int. Conf. on Information Technology Trends (ITT), 2020, pp. 1-6. IEEE.

[20] S. M. Mohammad and P. D. Turney, "NRC emotion lexicon," National Research Council, Canada, vol. 2, p. 234, 2013.

[21] S. Kiritchenko, and S. M. Mohammad, "Capturing nuances of sentiment and emotion in text." In Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval 2016) (pp. 446-455).

[22] Pixabay, "Pixabay," 2024. [Online]. Available: https://pixabay.com. [Accessed: July 4, 2024].