EMOLight: Immersive Visual Experience for the Audibly Impaired

***Abstract*— In this paper, we introduce EMOLight, an AI-driven ambient lighting system that enhances viewer immersion by synchronizing with the emotional content of audio cues. Leveraging Google's YAMNet, it performs real-time emotion recognition by analyzing audio cues. It applies Plutchik's emotion-color theory to dynamically adjust the ambient lighting, thus mirroring the emotional undertones of the audio. This synchronization 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, revolutionizing the way we experience and interact with media.**

Keywords—Ambient Lighting, 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 an advanced AI-driven solution that extends the concept of ambient lighting to include auditory elements. Our system utilises machine learning algorithms to analyse and classify audio cues from soundtracks, scores, and dialogues in real time. By integrating Robert Plutchik's emotion-colour theory [3], shown in Figure 1, our model dynamically adjusts the ambient lighting to reflect the emotional tone conveyed by the 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. Furthermore, concerted efforts have been undertaken to automatically annotate affective lexicons with nuanced intensity values using the WordNet Similarity software package [8].

Figure 1. The Plutchik's Emotion Wheel [9]

Crucially, our approach recognises that emotional responses to colour are subjective and can vary significantly between individuals and vary with place and culture [10]. As such, we incorporate user-specific adjustments, allowing the emotion-colour mappings to be tailored according to individual preferences and perceptions. This customization is particularly aimed at enhancing accessibility for the audibly impaired, providing them with an alternative sensory channel to perceive and engage with the emotional undertones of audiovisual content. Our project seeks to advance the state of the art in immersive visual and television technologies by seamlessly integrating audio-induced emotional cues into ambient lighting systems. Through this innovation, we aim to create a more inclusive and emotionally engaging viewing experience that can adapt to the diverse sensory needs and preferences of a broad

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

audience. This work contributes to the emerging field of multisensory enhancement technologies, which are increasingly important in the design of accessible and user-centric media systems.

II. Research Methodology

The research methodology of this project began with a systematic review of existing pre-trained machine-learning models, with a particular focus on YAMNet [11] due to its demonstrated proficiency in audio feature classification. The YAMNet deep learning model has demonstrated versatility and effectiveness across a variety of applications. For example, 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 [12]. Beyond emergency response, YAMNet has also been adapted for estimating human activites such as 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 [13].

Concurrently, we reviewed academic literature on music genre classification, analysing studies that utilized both well-known public datasets like GTZAN [14] and specialized private collections. This dual approach provided a comprehensive understanding of the state-of-the-art in audio classification and the 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.

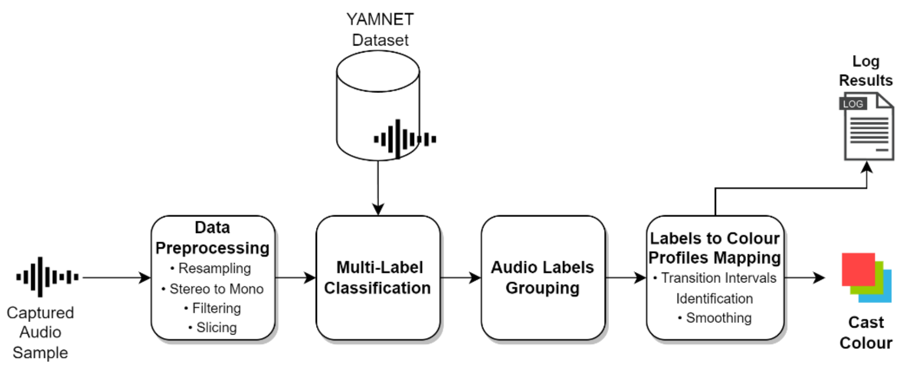


Figure.2 Overview of EMOLight

Our solution architecture involves rigorous preprocessing of continuously captured audio samples. 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 efficiently classified audio cues and assigned multi-labels based on the audio features detected. A significant challenge addressed in our project was the complexity of emotional responses to music, as depicted in Plutchik's emotion wheel. Our solution needed to handle the ambiguity inherent in musical emotion, where a single piece can evoke multiple, overlapping emotional responses. To tackle this, we developed a novel approach to effectively translate audio labels detected by YAMNet into associated emotional colours. This process utilized multiple human opinions to introduce variance, recognizing the subjectivity involved in interpreting audio cues into colour responses. Additionally, OpenAI's ChatGPT was employed to reduce bias and interpret subtle differences in word choice, essential for distinguishing between similar emotional cues. The NRC Word-Emotion Association Lexicon (EmoLex) [15] was also considered to provide a structured emotional association. The consensus of these sources is then considered by applying the mixture-of-experts approach. A sample of these audio labels and the defined consensus labels are shown in Table 1 below.

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 INITIAL RESULTS

Our initial testing phase involved using 10 royalty-free songs from Pixabay [16], spanning various genres and artists, to evaluate the accuracy of the proposed solution in identifying emotions. Each song was first analysed to understand the emotions it evoked, after which a 'main colour' representing the dominant emotion was assigned for quick visual comparison. This comparison method aimed to provide a reproducible and qualitative assessment of the program's performance in matching perceived emotions with corresponding colours. Table II below shows a few samples and their results.

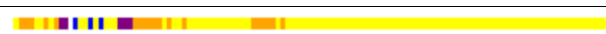
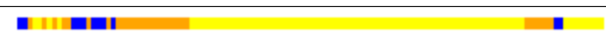
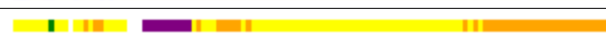
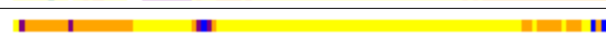
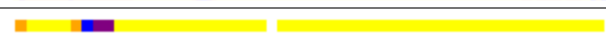
TABLE II. EMOLIGHT TESTING RESULTS – PIXABAY SONGS

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 also tested EMOLight's performance by analysing the 1967 film, *The Jungle Book*. This film was chosen for its diverse musical scores and range of emotions, providing a rich dataset for exploring EMOLight's capabilities. The film was segmented into eleven sections, each approximately ten minutes in duration, to facilitate data legibility and interpretation. This segmentation allowed for a detailed analysis of emotional shifts within the film, capturing the nuances of the narrative.

As detailed in Table III, the analysis revealed significant emotional variations across the film. Joyful events, such as Mowgli's playful interactions with Baloo, were predominantly displayed in yellow. In contrast, encounters with dangerous animals, like Shere Khan, were characterized by purple hues, reflecting fear and tension. Sad moments, such as Mowgli's separation from his wolf family, were depicted in blue, conveying a sense of loss and sorrow. Notably, white durations represent segments of the film where there is no background music or effects but only speech dialogues. This finding highlights the potential for further development of EMOLight to incorporate the emotional nuances of spoken language. This case study 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 analyzing the emotional impact of audio in film and other media.

TABLE III. EMOLIGHT TESTING RESULTS – THE JUNGLE BOOK FILM OF1967

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	

We further explored the application of our multi-label classification model by highlighting multiple colors based on the dominant colors identified. The results, as depicted in Figure 3, illustrate how multiple emotions can be detected in a test sample, indicating a potential mixture of emotions, with the most dominant color/emotion receiving greater emphasis. To enhance the representation of these mixed emotions or colors, several approaches could be considered. 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.

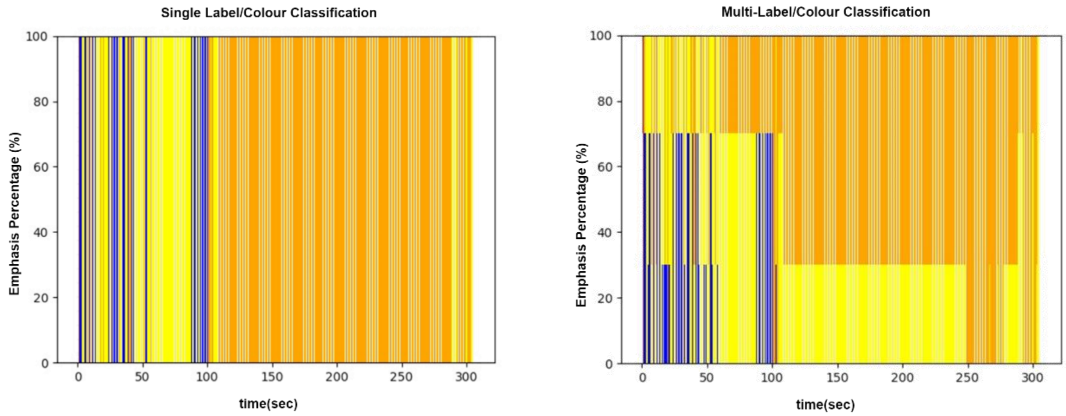


Figure.3 Single Label vs. Multi-Label/Colour Classification

IV. Conclusions

This research demonstrates the feasibility and potential of EMOLight, an AI-driven ambient lighting system that enhances viewer immersion and accessibility by synchronizing with the emotional content of audio cues. Leveraging machine learning and Plutchik's emotion-color theory, EMOLight provides real-time audio emotion recognition, user-specific customization, and multi-label classification for a personalized and engaging experience. Future research will focus on expanding emotion recognition capabilities, investigating advanced visualization techniques, conducting user studies, and exploring integration with other media formats. EMOLight paves the way for a future where technology adapts to diverse sensory needs and preferences, revolutionizing the way we experience and interact with media. By advancing multisensory enhancement technologies, our work sets new standards for accessible, engaging media technology and paves the way for further research in human-computer sensory interaction.

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. 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.

[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, S. S. Gokhale, "Mining Emotions on Plutchik's Wheel," 2020. [Online]. Available: https://doi.org/10.1109/SNAMS52053.2020.9336534

[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] 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.

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

[10] 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.

[11] M. Plakal and D. Ellis, "Yamnet," July 2024. [Online]. Available: https://github.com/tensorflow/models/tree/master/research/audioset/yamnet

[12] 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.

[13] 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.

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

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

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