Physiological Signals as Implicit Multimodal Input in Human-AI Interactions During Creative Tasks

Teodora Mitrevska
LMU Munich
Munich, Germany
teodora.mitrevska@ifi.lmu.de

Sven Mayer
LMU Munich
Munich, Germany
info@sven-mayer.com

ABSTRACT

Recent experiments have explored AI's role in creative collaboration, producing promising outcomes. While AI excels in generating text and visual art, human-AI collaboration lacks rich feedback compared to human-human interaction. So far, textual feedback is the dominant communication between systems and humans, which is in strong contrast to feedback in human-human conversations with many nonverbal cues. For creative individuals, their mental state, especially emotions, play a big role in creating art. They project their feelings on canvas or translate them into music. However, not being able to express them verbally puts a strain on the feedback and impacts the output of the collaboration. Using methods for emotion recognition can provide insights into humans' emotional states, which can be used to improve human-AI collaborations in creative scenarios. In this work, we present the concept of factoring in human physiological signals in co-creative scenarios with Generative AI as a multimodal input. We underscore this vision by presenting three applications utilizing an emotional-aware input.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Human Computer Interaction, physiological signals, generative AI

ACM Reference Format:
Teodora Mitrevska and Sven Mayer. 2024. Physiological Signals as Implicit Multimodal Input in Human-AI Interactions During Creative Tasks. In *Proceedings of GenAICHI: CHI 2024 Workshop on Generative AI and HCI*. ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

The latest capabilities of Generative AI allow users to create realistic imagery [17], music [1], and text [21], giving users the opportunity to be creative. Today, communication in the human-AI collaboration is based on text prompts; where the human prompts the AI to shape the output actively. Here, the AI collaborator solely relies on the text input in combination with the prompt history. We witness that the more detailed the prompt, the more control lies in the human's hands to steer the output. Still, this starkly contrasts with human-human conversations where humans supplement with many cues, such as tone of voice and nonverbal cues. However, today's prompts are usually required in text form, i.e., text-to-image, text-to-music, and text-to-text generation. Thus, the human input relies completely on their written proficiency. However, language is much more than just words, and expressing the feeling one might have while uttering words or drawing a picture is incredibly hard or even impossible. This limits the effectiveness of human-AI collaboration in creative tasks. Thus, while models are being perfected to deliver astonishing output based on text input, users lack the possibility to express themselves beyond text and sometimes images.

In the future, we envision multimodal input as an implicit way to steer the human-AI collaboration. In this work, we investigate which additional input methods can be considered in collaborative creative processes. In detail, we focus on physiological signals to enrich the input with the users' context. In this case, context refers to the state of the user body, e.g., workload [18] and visual attention [4]. Different physiological signals rapidly develop, allowing us to track and evaluate humans' mental states. So far, these technologies have been mostly used in the well-being context, providing insights into what signals our body exhibits. On a higher abstraction level, research has shown a correlation between physiological signals and the users' emotions, e.g., [10, 19]. Moreover, Mastria et al. [15] showed that emotional states can influence the creativity evaluation of ideas. Lubart [14] proposes that emotions play a central role in creative thinking. Finally, Dietrich [8] goes further by pointing towards a direct link between creativity and brain activity. Thus, we argue that co-creating with AI systems that are attuned to our state, using physiological signals, can foster deeper relational bonds between humans and the system.

When envisioning multimodal input using physiological signals to support human-AI collaboration, we raise several questions that future work must investigate in more depth. For instance, will it be possible that with such a multimodal input, we will receive more original and diverse output? Will this result in a stronger feeling of ownership? Will including our mental states in the communication with AI in creative processes give a new dimension to collaboration, fostering individual styles, differences, and higher originality?

Humans have been creating for centuries and have developed many methods to express their creativity. Some of the traditional methods included instruments to play music, paintbrushes to create art on canvas, and a quill to write poetry with. While the expressive arts have stayed, the methods of creation have developed further and provoked the formation of new creative directions and genres. While creating digital paintings includes quite some manual labor against the cost of autonomy, new and more sophisticated methods for art creation arise through Generative AI. Thus, in the following, we intend to lay the groundwork to understand the potential

GenAICHI: CHI 2024 Workshop on Generative AI and HCI

of physiological-aware multimodal human-AI collaboration. For this, we first look at the general feasibility and sketch out three showcases of such systems to support creative tasks.

2 PHYSIOLOGICAL SIGNALS AND APPLICATION

Webster's dictionary defines emotion as a conscious mental reaction (such as anger or fear) subjectively experienced as a strong feeling usually directed toward a specific object and typically accompanied by physiological and behavioral changes in the body. This definition triggers a multimodal approach in emotion recognition, meaning that emotions can reflect in multiple metrics [9]. We often communicate our feelings with body and face gestures, anxiety is exhibited in our sweat glands, and the production of theta waves hints at a deeply relaxed state. But in many cases, we experience "groups of emotions" rather than one single emotion [22]. Although emotion recognition can be quite challenging, it is definitely worth pursuing. There is a certain uniqueness in which emotions do we produce as a response to a particular stimuli and how large their intensity is. This data is strongly linked to identity, which is a core part of one's creativity. As already mentioned, emotion recognition requires a multimodal approach. However, some core methods can be extended or combined to work together, depending on the goal.

Electroencephalography (EEG). EEG is a noninvasive technique for the recording of electrical activity arising from the human brain on the human skull [3]. Moreover, Dietrich [8] linked EEG directly to creativity.

Electrocardiography (ECG). ECG is used as the conventional method for noninvasive interpretation of the electrical activity of the heart in real-time [9]. ECG can also be used for emotion recognition [9]. BIOPAC System MP150 can distinguish between joy and sadness [16], and Wireless bio sensor RF-ECG was used to measure negative arousal categorized into sadness, anger, disgust, and fear [12].

Electrodermal activity (EDA). EDA is a measure of neurally mediated effects on sweat gland permeability, observed as changes in the resistance of the skin to a small electrical current, or as differences in the electrical potential between different parts of the skin [7]. EDA is a particularly effective method to measure arousal. Namely, the armband from Bodymedia and Biosemi Active II device can offer good insights into valence and the level of arousal [5, 23]. Moreover, Chiossi et al. [6] adapted the environment using EDA and, thus, supported users in their goals, showcasing that EDA can support the user.

Other popular physiological signals that can be measured and input to the system as indicators of human emotions include skin temperature, blood volume pulse, electromyography, and respiratory signal [9, 20]. Although these methods require equipment to measure the signals, they have come far from requiring a controlled environment and complex devices. A literature review on measuring emotion conducted by Ba and Hu [2] implies that a total of 30 types of emotions were identified using wearables in past experiments. Some of them include: "happy/joy," "peacefulness," "anxiety," "anger" and "sadness."

Aside from their review being done in the education field, other fields debate the advantage of emotion recognition. Several computer hardware manufacturers plan to embed some physiological sensors into the control devices for video games, claiming that automatic emotion recognition is a key to creating affect-aware software. A prototype emotion recognition system proved that real-time emotion recognition is possible, through fusion of many input channels [20]. Another example derives from neuromarketing, where implicit feedback via EEG is gathered from the user, on whether they like a certain product or not [24]. Further examples are derived from Human-Robot Interaction where Kothig et al. [13] argued the benefits of physiological measurements in developing better relationships and communication with social robots.

3 SHOWCASES

Creative Writing. There are multiple AI writing assistants available nowadays. They assume different roles, e.g., Wordtune[1] is a tool that rewrites and rephrases writing. Grammarly[2] offers assistance in writing novels, Cowriter[3] specializes in creating content while mimicking your style, and Sudowrite[4] claims to be an expert in writing screenplays. From co-writing, spell-checking, and summarizing to even creating marketing content, it is clear that these models are quite proficient in what they do. However, the main input method stays at being text only, which could hinder the author from co-creating due to insufficient language proficiency or a simple inability to express emotions in words. Emotional awareness could bring depth to the writing process, for example by influencing the in narrative, creating heartwarming scenes triggered by positive emotions and channeling the emotional state of the author into characters.

Image Creation. DALL E2[5], DeepAI[6], Midjourney[7] and Pixlr[8] are just some of the sophisticated systems for image generation that we have available today. While the initial prompts are textual, they offer some additional versatility of choice in styles and models (DeepAI), or, on a more granular level: aspect, style, lightning, and color (Pixlr). Another interesting example of a creative Human AI collaboration is by Bertrand Gondouin, who utilized a generational model [11] to colorize Pokémons from sketches[9]. However, there are a lot of variables that the physiological signals of the user could manipulate. For example, the color palette, style, and intensity could be determined by the user's mood and express their true selves better through their creative works.

Music Generation. While Generative AI systems for music offer some additional options to manipulate with the output, their main input method remains text. In VeedIO[10], besides the prompt, the user can select the general style of the sound and set it to "chill,"

[1] https://www.wordtune.com
[2] https://www.grammarly.com
[3] https://cowriter.org/login
[4] https://www.sudowrite.com
[5] https://openai.com/dall-e-2
[6] https://deepai.org/machine-learning-model/text2img
[7] https://www.midjourney.com/home
[8] https://pixlr.com/image-generator/
[9] https://twitter.com/bgondouin/status/818571935529377792
[10] https://www.veed.io/tools/ai-music-generator

"hip-hop," "dance," etc. Other systems like Beatoven AI[11] allow the user to create music with somewhat more input. Depending on the content they want to generate music for, you can choose one of three options: music for a video, podcast, or just a composition. The users can also choose the genre or regional sound for their composition. Both Loudly[12] and Junia AI[13] allow text only as input. Not factoring in emotions into the creative process can show a lack of the nuanced emotional depth that is usually found in music. Composers often draw from personal experiences and emotions to create music that resonates authentically with themselves and their audiences. Enriching Generative AI systems with emotional awareness could help translating joy into a triumphant melody or a melody that describes an emotional trip from one mental state to another and therefore infuse authenticity, reliability, and originality into the work.

4 FUTURE WORK

While co-creating with Generative AI can lead to astonishing results, we identify that expressing human emotion through this kind of art can be difficult since, in most cases, the medium is limited to written text. This problem is potentially negatively impacting the creative process, as systems do not receive the physiological state of humans. Thus, we outlined how physiological signals could support human-AI collaborations during creative tasks. Equipping systems with such capabilities fosters new research avenues: What novel creative direction will emerge from these kinds of collaborative scenarios? Will there be a need for innovative interaction devices? What preferences will humans develop? And how might their emotion manifest in their digital artistic expressions? These questions point towards a future where human emotions and technological advancements intertwine to unlock new possibilities of creativity and expression, enabling new art forms.

REFERENCES

[1] Andrea Agostinelli, Timo Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. 2023. MusicLM: Generating Music From Text. arXiv:2301.11325 [cs.SD] https://arxiv.org/abs/2301.11325

[2] Shen Ba and Xiao Hu. 2023. Measuring emotions in education using wearable devices: A systematic review. *Comput. Educ.* 200, C (jul 2023), 20 pages. https://doi.org/10.1016/j.compedu.2023.104797

[3] Jeffrey W. Britton, Lauren C. Frey, Jennifer L. Hopp, Pearce Korb, Mohamad Z. Koubeissi, William E. Lievens, Elia M. Pestana-Knight, and Erik K. St. Louis. 2016. Electroencephalography (EEG): An Introductory Text and Atlas of Normal and Abnormal Findings in Adults, Children, and Infants. (2016). http://europepmc.org/books/NBK390354

[4] Niko A. Busch and Rufin VanRullen. 2010. Spontaneous EEG oscillations reveal periodic sampling of visual attention. *Proceedings of the National Academy of Sciences* 107, 37 (2010), 16048–16053. https://doi.org/10.1073/pnas.1004801107 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1004801107

[5] Guillaume Chanel, Joep JM Kierkels, Mohammad Soleymani, and Thierry Pun. 2009. Short-term emotion assessment in a recall paradigm. *International Journal of Human-Computer Studies* 67, 8 (2009), 607–627.

[6] Francesco Chiossi, Yagiz Turgut, Robin Welsch, and Sven Mayer. 2023. Adapting Visual Complexity Based on Electrodermal Activity Improves Working Memory Performance in Virtual Reality. *Proc. ACM Hum.-Comput. Interact.* 7, MHCI, Article 196 (sep 2023), 26 pages. https://doi.org/10.1145/3604243

[7] Hugo Critchley and Yoko Nagai. 2013. *Electrodermal Activity (EDA).* Springer New York, New York, NY, 666–669. https://doi.org/10.1007/978-1-4419-1005-9_13

[8] Arne Dietrich. 2004. The cognitive neuroscience of creativity. *Psychonomic Bulletin & Review* 11, 6 (01 Dec 2004), 1011–1026. https://doi.org/10.3758/BF03196731

[9] Andrius Dzedzickis, Artūras Kaklauskas, and Vytautas Bucinskas. 2020. Human Emotion Recognition: Review of Sensors and Methods. *Sensors (Basel, Switzerland)* 20, 3 (January 2020), E592. https://doi.org/10.3390/s20030592

[10] Maria Egger, Matthias Ley, and Sten Hanke. 2019. Emotion Recognition from Physiological Signal Analysis: A Review. *Electronic Notes in Theoretical Computer Science* 343 (2019), 35–55. https://doi.org/10.1016/j.entcs.2019.04.009 The proceedings of AmI, the 2018 European Conference on Ambient Intelligence..

[11] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. 2017. Image-To-Image Translation With Conditional Adversarial Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*

[12] Cai Jing, Guangyuan Liu, and Min Hao. 2009. The research on emotion recognition from ECG signal. In *2009 international conference on information technology and computer science*, Vol. 1. IEEE, 497–500.

[13] Austin Kothig, John Muñoz, Sami Alperen Akgun, Alexander M. Aroyo, and Kerstin Dautenhahn. 2021. Connecting Humans and Robots Using Physiological Signals – Closing-the-Loop in HRI. In *2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN'21).* 735–742. https://doi.org/10.1109/RO-MAN50785.2021.9515383

[14] Isaac Lubart, Todd I. nd Getz. 1997. Emotion, Metaphor, and the Creative Process. *Creativity Research Journal* 10, 4 (1997), 285–301. https://doi.org/10.1207/s15326934crj1004_1

[15] Serena Mastria, Sergio Agnoli, and Giovanni Emanuele Corazza. 2019. How does emotion influence the creativity evaluation of exogenous alternative ideas? *PLOS ONE* 14, 7 (07 2019), 1–16. https://doi.org/10.1371/journal.pone.0219298

[16] Kanlaya Rattanyu, Michiko Ohkura, and Makoto Mizukawa. 2010. Emotion monitoring from physiological signals for service robots in the living space. In *ICCAS 2010.* IEEE, 580–583. https://doi.org/10.1109/ICCAS.2010.5669914

[17] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis With Latent Diffusion Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* 10684–10695. https://doi.org/10.1109/CVPR52688.2022.01042

[18] Cornelia Setz, Bert Arnrich, Johannes Schumm, Roberto La Marca, Gerhard Tröster, and Ulrike Ehlert. 2010. Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine* 14, 2 (2010), 410–417. https://doi.org/10.1109/TITB.2009.2036164

[19] Lin Shu, Jinyan Xie, Mingyue Yang, Ziyi Li, Zhenqi Li, Dan Liao, Xiangmin Xu, and Xinyi Yang. 2018. A Review of Emotion Recognition Using Physiological Signals. *Sensors* 18, 7 (2018). https://doi.org/10.3390/s18072074

[20] Mariusz Szwoch and Wioleta Szwoch. 2015. Emotion Recognition for Affect Aware Video Games. In *Image Processing & Communications Challenges 6*, Ryszard S. Choraś (Ed.). Springer International Publishing, Cham, 227–236.

[21] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv:2307.09288 [cs.CL]

[22] Debra Trampe, Jordi Quoidbach, and Maxime Taquet. 2015. Emotions in Everyday Life. *PLOS ONE* 10 (12 2015), 1–15. https://doi.org/10.1371/journal.pone.0145450

[23] Olivier Villon and Christine Lisetti. 2007. Toward recognizing individual's subjective emotion from physiological signals in practical application. In *Twentieth IEEE International Symposium on Computer-Based Medical Systems (CBMS'07).* IEEE, 357–362.

[24] Li Zeng, Mengsi Lin, Keyang Xiao, Jigan Wang, and Hui Zhou. 2022. Like/Dislike Prediction for Sport Shoes With Electroencephalography: An Application of Neuromarketing. *Frontiers in Human Neuroscience* 15 (01 2022). https://doi.org/10.3389/fnhum.2021.793952

[11] https://www.beatoven.ai/
[12] https://www.loudly.com/text-to-music
[13] https://www.junia.ai/tools/music-generator