Human Interaction and Emerging Technologies (IHIET 2026), Vol. 236, 2026, 341–347
https://doi.org/10.54941/ahfe1008109

Toward Culturally Sensitive Emotion Estimation Through Multimodal Sensing and Everyday Behavioral Cues

Meina Tawaki[1], Ichi Kanaya[1], Munenori Koyasu[2], and Keiko Yamamoto[2]

[1]Graduate School of Integrated Science and Technology, Nagasaki University, Nagasaki, 852-8521, Japan

[2]Department of Design Engineering and Technology, Tokyo Denki University, Tokyo, 120-8551, Japan

ABSTRACT

This study provides new insights into emotion estimation technologies by integrating theoretical model comparison, multimodal sensing, and proposes a conceptual framework toward multicultural analysis. First, we compare representative emotion models, including Russell's two-dimensional circumplex model and Plutchik's three-dimensional model, examining their theoretical foundations, expressive capacities, and applicability to computational systems. This comparison clarifies their strengths and limitations in practical emotion estimation. Based on this analysis, we propose an emotion estimation approach that combines physiological and behavioral data. Heart rate signals from wearable sensors are integrated with environmental pressure data, along with behavioral indicators such as chair-leaning posture and beverage consumption timing. This multimodal framework enables more comprehensive emotion estimation beyond physiological signals alone. We further analyze the relationship between estimated emotions and behavioral patterns. Statistical results show meaningful correlations between specific behaviors and emotional dimensions, demonstrating that behavioral data can complement physiological measurements. This integration improves robustness and interpretability, particularly in situations where physiological data are noisy or insufficient. Additionally, the study shows how it would adopt a multicultural perspective by examining how cultural differences influence emotional expression and physiological responses. The findings reveal that models developed in a single cultural context may not generalize well, emphasizing the importance of cultural adaptability in emotion estimation systems. Overall, this research advances affective computing by combining established models with multimodal sensing and cultural considerations. The proposed framework supports more accurate, inclusive, and practical emotion-aware systems, with applications in human–computer interaction, social robotics, and adaptive technologies.

Keywords: Emotion estimation, Multimodal sensing, Behavioral cues, Cultural differences, Affective computing, Human–computer interaction

INTRODUCTION

In recent years, digital devices are increasingly required not only to provide information but also to adapt their behavior according to users' emotional

Received June 11, 2026; Revised July 10, 2026; Accepted August 1, 2026; Available online August 25, 2026

states. This concept is a central topic in Affective Computing (Picard, 1997), which aims to recognize and respond to human emotions.

However, conventional emotion estimation methods mainly rely on physiological signals (Koelstra et al., 2012) such as heart rate or brain activity. While these signals can provide reliable information, they are often intrusive and difficult to apply in everyday situations.

Therefore, this study focuses on incorporating daily behavioral data into emotion estimation. The objective is to develop a more natural and non-intrusive approach by integrating physiological data with environmental and behavioral data. Emotion is evaluated using Russell's two-dimensional model consisting of Pleasure–Displeasure and Arousal–Sleepiness (Mehrabian, 1996).

RELATED WORK AND RESEARCH GAP

Various emotion models have been proposed in previous research, including Ekman's basic emotions (Ekman, 1992) and Plutchik's model (Plutchik, 1980). Among them, Russell's circumplex model is widely used because it represents complex emotions using two simple dimensions (Russell, 1985) (Figure 1).

Based on Russell's circular model, Kato et al. developed a new emotion estimation model. Focusing on the relationship between emotions and physiological processes, they defined the axes and mapped emotions based on evidence from medical literature and other sources. They identified electroencephalogram (EEG) and heart rate waveforms—whose relationship to emotions has been medically demonstrated—as biological signals suitable for emotion estimation. Since EEG can be used to quantify arousal levels and heart rate waveforms can be used to quantify the activation levels of the sympathetic and parasympathetic nervous systems, the vertical and horizontal axes were defined as shown in Figure 2. The correlations between emotions and arousal, as well as between emotions and the activation levels of the sympathetic and parasympathetic nervous systems, were organized, and the emotions to be placed in the four quadrants were determined. The performance verification results demonstrated that applying a model based on medical facts regarding the effects of emotions on the organs underlying the physiological signals has the potential to significantly improve performance compared to conventional models (Kato, 2023).

Existing emotion estimation approaches mainly rely on:

- Facial expression analysis (Pantic & Rothkrantz, 2000)
- Physiological signals such as heart rate and EEG

However, these approaches have several limitations. Facial expressions are affected by individual differences and may be consciously controlled, while physiological measurements often require specialized equipment that limits daily usability.

Recent studies have explored behavioral data, such as posture and movement, but few have integrated both physiological and behavioral signals.

Is emotion primarily reflected in internal physiological states, or can it be captured through external behaviors?

This study addresses this question by proposing a multimodal approach that combines both aspects.

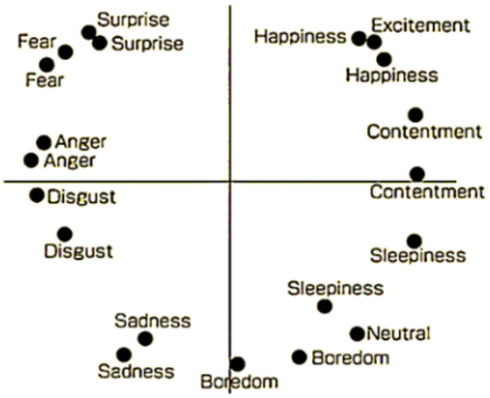

Figure 1: Examples of adult mental space (Russell & Bullock,1985).

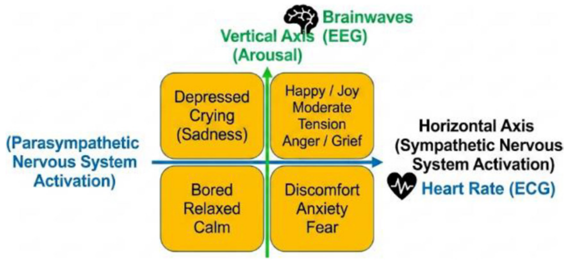

Figure 2: The emotion estimation model (Kato, 2023).

PROPOSED METHOD

This study integrates three types of data for emotion estimation:

- Physiological data: heart rate (BPM), HRV, RMSSD
- Behavioral/environmental data: chair leaning pressure, drinking tea actions (cup lifting)
- Subjective data: emotion logger (Pleasure and Arousal sliders).

These data represent three distinct perspectives:

Internal state (physiology), External behavior (actions), Subjective perception (self-report)

To analyze these data, statistical methods including correlation analysis, ANOVA, multiple regression analysis, and principal component analysis (PCA) were applied.

Emotion is treated not as a single measurable variable but as a combination of multiple weak signals.

Figure 3: Cup pressure sensor.

EXPERIMENTAL METHOD

An experiment was conducted with nine university students. Three video stimuli were used to induce different emotional states:

- Movie trailer: high arousal and positive valence
- Centipede video: high arousal and negative valence
- Bonfire video: low arousal and positive valence.

During video viewing, sensor data were continuously recorded, and participants reported their emotions using a slider-based emotion logger. Drinking tea behavior was not restricted in order to allow natural actions.

This experimental design emphasizes ecological validity by approximating daily-life conditions rather than tightly controlled laboratory conditions.

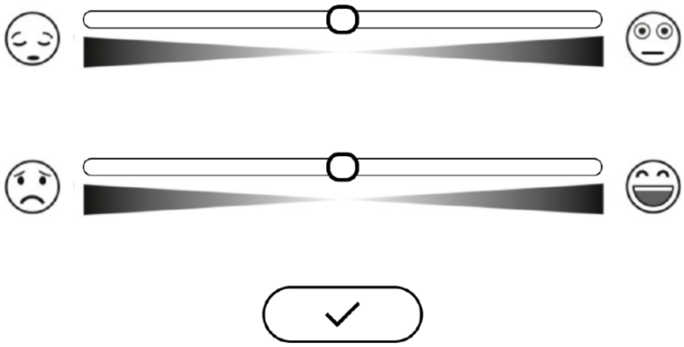

Figure 4: Slider-style emotion logger.

RESULTS

The emotion logger results confirmed that the intended emotional states were successfully induced by each video. The subjective evaluations were consistent with the distribution of Russell's emotion model.

For physiological data, weak positive correlations were observed between HRV/RMSSD and Pleasure, while MeanHR showed little relationship.

No statistically significant differences among video types were found in the ANOVA analysis.

Behavioral data showed that drinking tea frequency increased after the unpleasant (centipede) video, suggesting that stress or discomfort may influence behavior.

However, the integrated regression model showed low explanatory power (low R^2), and PCA did not reveal clear clusters of emotional states.

Only weak relationships were observed between sensor data and emotions.

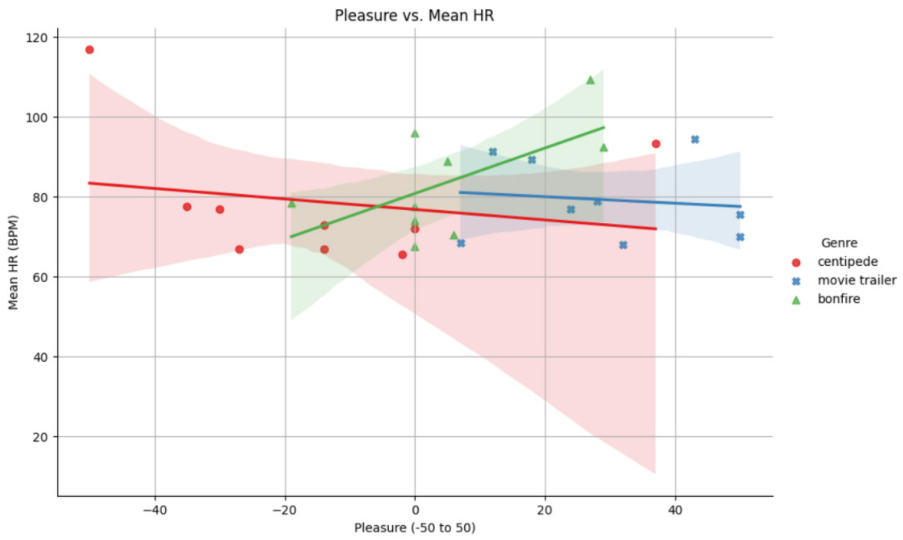

Figure 5: Average heart rate and emotion logger.

DISCUSSION

The empirical results of this baseline study indicate that human emotion cannot be fully explained (Scherer, 2005) or captured by a single type of data. Both physiological metrics and everyday behavioral data partially reflect internal emotional states, but their individual contributions remain limited when isolated. Importantly, this research revealed the clear presence of multiple weak correlations across the distinct data streams. While each signal alone is insufficient for deterministic modeling, their systematic combination can provide meaningful, nuanced information for affective computing systems.

Based on these findings, we argue that emotion should be conceptually understood as:

- **Multimodal**, requiring a combination of completely different internal and external signals.
- **Probabilistic**, meaning it is inherently non-deterministic and cannot be perfectly predicted via rigid rules.
- **Trend-based**, favoring long-term behavioral and physiological drift rather than discrete classification.

Therefore, the low initial prediction accuracy observed in our integrated regression models should not be interpreted as a failure of the methodology. Instead, it serves as empirical evidence of the sheer complexity of human emotions when evaluated under ecologically valid, everyday conditions.

Establishing this specific multimodal framework—which treats emotion as an aggregate of multiple weak signals—is a critical step toward developing truly culturally adaptive emotion estimation systems. Prior research emphasizes that emotion models developed within a single cultural context often fail to generalize well across diverse populations. This limitation exists because cultural differences significantly influence both outward emotional expressions and internal physiological responses.

The Vision for Cultural Adaptability: Traditional affective computing relies heavily on strong, direct indicators like facial expressions, which are easily controlled or altered by cultural display rules. By shifting the focus toward a probabilistic combination of everyday behavioral cues (such as chair-leaning and beverage consumption timing) alongside physiology, our framework introduces a highly flexible architecture. Rather than deploying a rigid, monocultural classification standard, this weak-signal integration can eventually be calibrated to accommodate the unique behavioral baselines and emotional trends of different cultural groups.

FUTURE WORK

Acknowledgeable limitations of this preliminary baseline study include:

- **Small Sample Size:** The initial evaluation was restricted to a small cohort of nine university students.
- **Limited Feature Design:** The number of environmental and behavioral features extracted from the sensor data was narrow.
- **Subjective Variability:** There was high individual variability in how participants utilized the slider-based subjective emotion loggers.

To transition this foundational framework from a localized baseline into a culturally sensitive emotion estimation system, future research will pursue the following trajectories:

- **Dataset Expansion:** We aim to significantly increase the dataset size by intentionally recruiting a multicultural cohort of participants to analyze cross-cultural behavioral variances.
- **Advanced Machine Learning:** We plan to improve feature extraction and apply machine learning and deep learning architectures capable of mapping complex, non-linear cultural trends.
- **Long-Term Real-World Deployment:** Future iterations will move beyond video stimuli to conduct long-term, real-world measurements, evaluating how daily behavioral cues naturally shift across different cultural environments.

CONCLUSION

This study demonstrated the potential of integrating behavioral data with physiological signals for emotion estimation. Although prediction accuracy remains limited, the results suggest a promising direction toward **non-intrusive emotion estimation based on daily behavior**.

ACKNOWLEDGMENT

This work was supported by JST SPRING, Japan Grant Number JPMJSP2172.

REFERENCES

Darwin, Charles. (1872) The expression of the emotions in man and animals. London: John Murray.

Diener, Ed. Emmons, Robert A. (1984) The independence of positive and negative affect. *Journal of Personality and Social Psychology*.

Ekman, Paul. (1992) An argument for basic emotions. *Cognition & Emotion*.

Kato, Takumi. (2023) Emotion estimation model based on physiological signals and medical evidence.

Koelstra, Sander. Mühl, Christian. Soleymani, Mohammad et al. (2012) DEAP: A database for emotion analysis using physiological signals. *IEEE Transactions on Affective Computing*.

McDuff, Daniel. el Kaliouby, Rana. Picard, Rosalind W. (2012) Crowdsourcing facial responses to online videos. *IEEE Transactions on Affective Computing*.

Mehrabian, Albert. (1996) Pleasure-arousal-dominance: A general framework for describing temperament. *Current Psychology*.

Pantic, Maja. Rothkrantz, Leon J. M. (2000) Automatic analysis of facial expressions: The state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Picard, Rosalind W. (1997) *Affective computing*. Cambridge, MA: MIT Press.

Plutchik, Robert. (1980) *Emotion: A psychoevolutionary synthesis*. New York: Harper & Row.

Russell, James A. (1980) A circumplex model of affect. *Journal of Personality and Social Psychology*.

Russell, James A. Bullock, Merry. (1985) Multidimensional scaling of emotional facial expressions. *Journal of Personality and Social Psychology*.

Scherer, Klaus R. (2005) What are emotions and how can they be measured. *Social Science Information*.

Schuller, Björn W. Steidl, Stefan. Batliner, Anton. (2011) The INTERSPEECH 2011 speaker state challenge. *INTERSPEECH Proceedings*.