From Recognition to Expression: A Qualitative Study on Emotion-Driven Mechanisms in Interactive Art

Yiming He[a]
Interactive Art, China Agricultural University, Haidian, Beijing, China

Keywords: Interactive Art, Emotion Recognition, Affective Computing, Feedback Mechanism, Multimodal Design.

Abstract: As emotion recognition technologies evolve, their integration into interactive art opens up new possibilities for real-time, affect-sensitive experiences. Rather than relying on conventional inputs like touch or movement, emotion-driven systems detect users' affective states—through facial expressions, vocal features, or physiological signals—and translate them into dynamic visual, auditory, or spatial feedback. This approach not only creates novel aesthetic interactions but also deepens emotional resonance between participant and system. This paper examines the core mechanisms behind emotion-driven interaction in art by analyzing how emotional data is captured, interpreted, and expressed. It explores three major recognition pathways—facial, vocal, and physiological—each offering distinct affordances in responsiveness, sensitivity, and ambiguity. The paper further investigates how artists translate raw emotion into aesthetic parameters, embedding their own conceptual logic and cultural framing into the system's response. Through comparative case analysis and cross-modal reflection, this study highlights emotion's dual role as both computational input and expressive medium, and proposes future directions for designing emotionally intelligent, adaptive, and culturally sensitive interactive systems.

1 INTRODUCTION

Emotion is one of the core dimensions of human experience and a key trigger in aesthetic perception (Picard, 1997; Norman, 2004). With the evolution of affective computing and the advent of real-time recognition technologies, a new genre of interactive art has emerged—emotion-driven systems where user input is not given via buttons or gestures, but rather through affective states. These systems capture emotion from facial expressions, vocal tone and biometric signals, and respond with changes in visuals, sound, or spatial dynamics, forming a loop of recognition and expression.

Such emotion-based interaction not only transforms the traditional author-audience dynamic but also challenges the structure of feedback in design. Here, the participant becomes both sender and receiver, and the work becomes a living mirror reflecting shifting internal states. Rather than outputting fixed responses, these systems embed variability, ambiguity, and aesthetic responsiveness into their logic—prompting a reconsideration of agency, authorship, and perception in computational media art (Löwgren & Stolterman, 2004).

This paper aims to map out the emotion-driven mechanism in interactive art by examining three core aspects: the algorithmic methods for recognizing and categorizing emotional states; the logic through which affective variables are translated into aesthetic output; and the ways different modalities—facial, vocal, and physiological—shape the structure and sensitivity of feedback loops. By comparing systems across recognition modes and analyzing their technical and experiential characteristics, this paper aims to provide a theoretical and practical foundation for future design in emotion-aware interaction art.

2 EMOTION RECOGNITION TECHNOLOGY AND CLASSIFICATION

Emotion recognition technology lies at the foundation of affect-driven interactive systems. In contemporary

[a] https://orcid.org/0009-0002-7163-2091

He, Y.
From Recognition to Expression: A Qualitative Study on Emotion-Driven Mechanisms in Interactive Art.
DOI: 10.5220/0014355900004718

In *Proceedings of the 2nd International Conference on Engineering Management, Information Technology and Intelligence (EMITI 2025)*, pages 376-380
ISBN: 978-989-758-792-4

practice, three major recognition pathways have become prominent: facial expression analysis, vocal feature extraction, and physiological signal detection. Each operates under different technical paradigms and offers distinct advantages and constraints, shaping both the immediacy and depth of emotional feedback in artistic settings.

2.1 Facial Emotion Recognition

Facial expression recognition is one of the most widely used approaches in emotion-aware interaction. Building upon Ekman and Friesen's (1978) Facial Action Coding System (FACS), current systems utilize convolutional neural networks (CNNs) to extract key features such as eyebrow movement, eye openness, and mouth curvature (Ekman et al.,1978). These features are then classified into discrete emotional categories—typically six or seven "basic emotions" (e.g., happiness, anger, sadness, fear).

Compared to other recognition types, facial analysis is visual and intuitive, making it ideal for immediate, image-based feedback. However, this method is often limited by cultural variation, individual expressiveness, and external environmental factors such as lighting or occlusion (Jack et al., 2012). Furthermore, some scholars argue that emotions expressed on the face are often socially regulated and do not always correspond to internal states (Barrett et al., 2019), posing challenges to the validity of such recognition systems.

2.2 Vocal Emotion Recognition

Vocal emotion recognition analyzes non-verbal prosodic features of speech—such as pitch, volume, tempo, and pause duration—to infer affective states. Using support vector machines (SVMs), hidden Markov models (HMMs), or more recently deep learning frameworks (e.g., LSTM networks), systems can detect emotional shifts in both scripted and spontaneous speech (Schuller et al., 2011).

One of the advantages of this modality is that it captures temporal dynamics, allowing artists to create feedback that evolves with the flow of audience speech. However, its performance is susceptible to background noise, language differences, and speaker variability (Eyben et al., 2016). Moreover, the ambiguity of tone and context in a human voice often complicates the classification process, necessitating hybrid approaches that integrate semantic and acoustic cues.

2.3 Physiological Emotion Recognition

Physiological recognition focuses on collecting biometric signals such as heart rate variability (HRV), galvanic skin response (GSR), and brainwave patterns (EEG). These signals, often captured via wearables or sensors, offer a more embodied and involuntary indication of affective states (Kim et al., 2004). Since physiological responses are harder to consciously manipulate, this modality is generally considered more reliable for detecting subtle or implicit emotional changes.

Nonetheless, physiological systems face challenges in terms of signal stability, device comfort, and individual baseline differences. Emotional interpretation from such signals often requires complex, personalized calibration and machine learning models (Zhao et al., 2017). Despite this, physiological recognition opens up compelling avenues for biofeedback art, where internal states are transformed into visual, auditory, or spatial experiences.

3 TRANSLATION MECHANISMS: FROM RECOGNITION TO EXPRESSION

The process of translating emotion into artistic output is not merely a matter of data conversion, but a complex operation involving semantic interpretation, aesthetic decisions, and perceptual synchronization. In emotion-driven interactive systems, recognition results—often numerical or categorical—must be transformed into parameters that shape dynamic feedback, such as image color, sound tempo, or spatial arrangement. This section investigates the logic and strategies that guide this mapping process.

3.1 Variable Mapping and Modality Coupling

At the heart of the translation mechanism lies the mapping between affective variables and visual or auditory elements. For instance, an increase in "valence" (positive emotional intensity) may lead to warmer color palettes and smoother visual transitions, while higher "arousal" may be expressed through faster animations or sharper sonic pulses (Russell, 1980). Such mappings rely on semiotic associations derived from psychology and media theory: red for passion, blue for calmness, acceleration for tension, etc.

In multimodal systems, coupling between modalities requires synchronization. A user's excited tone of voice might correspond to both a brightening visual field and an increase in ambient sound volume. Artists thus play a crucial role in designing these mappings—what Schubert (2001) refers to as "aesthetic translation frameworks." The decision of whether sadness manifests as grayscale visuals or low-frequency drones is not fixed but authored, shaped by the artist's conceptual priorities and expressive intentions (Schubert, 2001).

3.2 Real-Time Responsiveness and Adaptive Logic

Emotion-driven feedback systems often operate under real-time constraints, requiring that translation occurs within milliseconds to maintain the illusion of immediate response. This poses challenges in computational load, latency management, and interpretive ambiguity. To address this, many systems implement threshold-based logic or fuzzy classification, ensuring that emotional inputs are translated into feedback that is perceptible but not erratic (El Ayadi et al., 2011).

Some advanced systems incorporate adaptive algorithms that learn from user behavior, gradually adjusting translation rules to better match individual expressivity. These systems reflect a shift from static mappings to relational models, where meaning is co-constructed through repeated interaction (Höök, 2008). This marks a move toward "affective adaptivity," in which the system not only responds to emotion but evolves with it.

3.3 Cultural Semiotics and Subjective Interpretation

Although mapping logic is often implemented through computational models, its meaning is shaped by cultural codes and audience interpretation. The same musical cue may evoke joy in one cultural context and nostalgia or melancholy in another (Matsumoto, 1990). Thus, emotional translation must balance between universal affective principles and local cultural semantics.

Artists working with global audiences often adopt hybrid strategies—allowing participants to calibrate feedback themselves, or designing intentionally ambiguous outputs that invite open-ended interpretation. This aligns with principles in critical interaction design, where the goal is not to control perception but to create spaces for emotional reflection and relational meaning-making (Gaver et al., 2003).

4 FEEDBACK MECHANISMS IN EMOTION-DRIVEN SYSTEMS

If recognition and translation form the input logic of affective interaction, then feedback is where the system's output materializes—manifesting as light, sound, movement, or environment. In emotion-driven art, feedback is not just a response but a communicative gesture, often designed to mirror, amplify, or modulate the user's internal state. This section analyzes three typical types of feedback systems-facial-based, vocal-based, and physiological-based-to highlight their structural features, expressive dynamics, and design implications.

4.1 Facial-Based Feedback: Mirroring and Contrast

In facial emotion recognition systems, feedback is often visual and directly tied to the participant's expression. Some installations adopt a "mirroring" logic, projecting a facial image back to the user with augmented emotional cues—such as intensified smiles or exaggerated sadness. This technique reinforces emotional awareness and creates a feedback loop of self-perception (Yoo et al., 2012). For example, in Rafael Lozano-Hemmer's 33 Questions per Minute, faces captured by a camera trigger projected responses that vary in scale and intensity based on micro-expressions.

Other systems pursue contrast instead of mirroring-displaying opposite or ambiguous emotional feedback to encourage reflection or emotional disruption. Such techniques invite users to reconsider the reliability of their own affective projections and provoke deeper introspection (LaBelle, 2010). Whether mirroring or contrasting, facial-based systems emphasize the immediacy and recognizability of visual emotion, making them suitable for installations that prioritize facial presence and social feedback.

4.2 Vocal-Based Feedback: Rhythm and Tonality

Vocal-based systems often translate emotion into temporal feedback—modulating rhythm, pitch, or musical structure. One typical strategy is to map vocal arousal levels to audiovisual tempo: an excited voice may quicken light flashes or increase the pace of background sound. In Chikashi Miyama's Sonic Emotion Space, for example, the user's vocal tone controls ambient noise density and movement, creating a multi-sensory landscape that evolves with the emotional voice stream.

This form of feedback emphasizes the performative and musicality of emotion, allowing participants to "compose" a real-time affective environment through tone alone. However, such systems often struggle with subtle emotional cues and require careful calibration to avoid over-sensitivity or misclassification (Schröder et al., 2011). Nevertheless, they are particularly powerful in sound-based installations or immersive audio experiences.

4.3 Physiological-Based Feedback: Internal States Externalized

Physiological feedback systems transform invisible bodily signals into experiential elements. Heart rate may control visual pulsation, skin conductivity may affect color gradients, and EEG data may trigger ambient transitions. In Lisa Park's Eunoia, EEG readings are mapped to water vibration, turning mental concentration into a tangible aesthetic experience (Park, 2013).

Because physiological signals are often involuntary and less consciously mediated, their translation into feedback can create a sense of intimacy or vulnerability. Viewers not only "see" their inner states but must confront their opacity and volatility. Such works suggest that emotion-driven systems are not only responsive but reflective—inviting users to inhabit a loop between sensing and being sensed (Dourish, 2001).

5 COMMONALITIES AND DIFFERENCES: COMPARATIVE MECHANISMS AND AESTHETIC CHARACTERISTICS

5.1 Technical Structures and Mapping Clarity

All three systems aim to deliver real-time affective responses, although their technical structures differ. Facial and vocal mappings are often straightforward, relying on common emotional cues such as smiles, tone, or volume, while physiological responses are more subtle and ambiguous. For example, an increased heart rate could indicate excitement or anxiety. This ambiguity invites users to interpret and engage with feedback subjectively, turning emotional responses into reflective acts. Artists and designers must decide how much clarity or openness to embed in the mapping logic, balancing between direct impact and poetic ambiguity.

5.2 Cultural Perception and Subjective Framing

Emotional expression is shaped by cultural context. Facial cues such as anger or surprise may be universal to some extent, but their recognition varies across regions (Matsumoto, 1990). Vocal emotionality is especially affected by language and tone expectations, which influence how systems perceive and render feedback (Scherer, 2003). Designers must account for such variations to avoid misclassification or miscommunication. The same physiological input may have different emotional interpretations depending on cultural learning and symbolic associations, reinforcing the need for contextual sensitivity in system design.

5.3 Artistic Intention and Aesthetic Expression

Ultimately, the artist plays a critical role in composing the emotional narrative. From choosing which emotions are detectable, to designing how outputs should appear or evolve, the artist defines the system's expressive logic. Feedback is not merely a mechanical output, but an aesthetic decision—whether it mirrors, contrasts, or abstracts emotion—crafted to deepen user engagement. In this way, emotion becomes not just data, but an artistic medium: shaped, framed, and performed in the space between subject and system.and relational meaning-making (Gaver et al., 2003).

6 CONCLUSIONS

Emotion-driven interactive art reveals how emotional states can serve not only as data but as expressive media. Through a layered process of recognition, translation, and feedback, interactive systems are capable of reflecting, interpreting, and amplifying human effects in real time. These systems do not passively register emotion—they reshape it, choreograph it, and sometimes even challenge it. Artists play a pivotal role in authoring these emotional narratives, embedding ambiguity, rhythm, and cultural nuance into the logic of interaction.

By comparing multiple modalities—facial, vocal, and physiological—this paper has outlined the technological, aesthetic, and symbolic logic underlying affective interaction. Each modality offers distinct affordances, but they are unified by a common goal: to create a dynamic feedback loop between human emotion and artistic expression. In this loop, emotion is no longer a static input; it becomes performative, interpretive, and affectively resonant.

Looking forward, future research and creative work should emphasize not only the technical accuracy of emotion recognition but also the poetic potential of emotional ambiguity. Designers must consider cultural diversity, user agency, and the affective ethics of machine interpretation. Rather than narrowing emotion into rigid classifications, interactive art should aim to open new spaces for emotional experience—ones that are reflective, participatory, and deeply human. Ultimately, the true potential of emotion in interactive art lies not in control, but in connection.

REFERENCES

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. Perspectives on Psychological Science, 14(6), 917 – 933.

Dourish, P. (2001). Where the action is: The foundations of embodied interaction. MIT Press.

Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System (FACS). Consulting Psychologists Press.

El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. Speech Communication, 53(9 – 10), 1162 – 1181.

Eyben, F., Scherer, K. R., Schuller, B. W., Sundberg, J., André, E., Busso, C., ... & Truong, K. P. (2016). The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Transactions on Affective Computing, 7(2), 190 – 202.

Gaver, W. W., Beaver, J., & Benford, S. (2003). Ambiguity as a resource for design. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 233 – 240.

Höök, K. (2008). Affective loop experiences—What are they? Proceedings of the 3rd International Conference on Persuasive Technology, 1 – 12.

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2012). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. Proceedings of the National Academy of Sciences, 109(15), 7241 – 7246.

Kim, J., & André, E. (2004). Emotion recognition based on physiological changes in music listening. Proceedings of the 7th International Conference on Pattern Recognition (ICPR), 4, 400 – 403.

LaBelle, B. (2010). Acoustic territories: Sound culture and everyday life. Bloomsbury Publishing.

Löwgren, J., & Stolterman, E. (2004). Thoughtful interaction design: A design perspective on information technology. MIT Press.

Matsumoto, D. (1990). Cultural similarities and differences in display rules. Journal of Personality and Social Psychology, 58(1), 128 – 134.

Norman, D. A. (2004). Emotional design: Why we love (or hate) everyday things. Basic Books.

Park, L. (2013). Eunoia. http://lisaapark.com/eunoia

Picard, R. W. (1997). Affective computing. MIT Press.

Russell, J. A. (1980). A circumplex model of affect. Psychological Review, 87(2), 145 – 153.

Schröder, M., Devillers, L., Cowie, R., Douglas-Cowie, E., & Batliner, A. (2011). Approaches to emotion recognition in speech: Towards emotional corpora. IEEE Transactions on Affective Computing, 3(2), 132 – 144.

Schubert, E. (2001). Continuous measurement of self-report emotional response to music. Music Perception, 23(1), 27 – 46.

Schuller, B., Rigoll, G., & Lang, M. (2011). Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine – belief network architecture. IEEE Transactions on Affective Computing, 2(1), 32 – 45.

Yoo, Y., Jin, B., & Myung, R. (2012). Real-time feedback display for facial expression recognition using facial feedback. Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 689 – 690.

Zhao, M., Adib, F., & Katabi, D. (2017). Emotion recognition using wireless signals. IEEE Transactions on Affective Computing, 8(4), 439 – 451.