6 Exploring Affective Images and Emotion Elicitation in Embodied Emotions

Nataša Kovač, Faculty of Applied Sciences, University of Donja Gorica, Oktoih 1, Podgorica, 81000, Montenegro, https://orcid.org/ [0000-0002-6671-2938](https://orcid.org/0000-0002-6671-2938)

Hojjatollah Farahani, Department of Psychology, Faculty of Humanities, Tarbiat Modares University, Tehran, Iran, <https://orcid.org/0000-0002-9799-7008>

Peter Watson, Cognition and Brain Sciences Unit, University of Cambridge, 15 Chaucer Road, Cambridge, CB2 7EF, United Kingdom, https://orcid.org/0000-0002-9436-0693

**Abstract:** The chapter begins by illuminating the pivotal role of visual stimuli in understanding emotions, highlighting their significance within the context of embodied emotion. It then delves into the intricate relationship between facial expressions and affective imagery, unravelling the complex interactions between the human body and emotions within a visual context. The chapter also explores the application of machine learning approaches tailored for affective image analysis, providing a comprehensive understanding of how AI can be leveraged to decode emotional responses from visual data. We then examine the intricate process of embodied emotion elicitation, shedding light on its significance in understanding and recognising emotions. The chapter then unveils the concept of embodied emotion priming, a powerful technique for preparing individuals to experience specific emotions. Moving forward, we delve into the methods and mechanisms employed to provoke embodied emotions, offering insights into the strategies employed. Furthermore, we explore the quantification of the impact of priming and provocation techniques on embodied emotions, providing a framework for measurement. Lastly, we examine how emotion elicitation has evolved in the digital age, highlighting its implications for contemporary research and applications in the field of artificial psychology.

**Keywords:** Visual stimuli; Emotion elicitation; Computational modelling; Classification performance

# 6.1 Introduction

Human emotions do not exist in isolation; they manifest vividly through interactions between individuals and their environments. Central to these interactions are visual stimuli (faces, body postures, contextual scenes) which powerfully evoke and shape emotional experiences. The processes of emotion elicitation, particularly through visual and embodied modalities, have profound implications for understanding emotional dynamics. This chapter explores how affective images and deliberate emotion elicitation techniques help uncover the intricate connections between visual perception, bodily responses, and emotional experience. By integrating historical insights, contemporary methodologies, computational advances, and ethical considerations, we establish a comprehensive understanding of how visual and emotional stimuli shape embodied emotional responses.

In recent decades, emotion research has progressively integrated the concept of embodiment, recognising that affective experience is deeply rooted in bodily processes that interact with neural and perceptual systems. This embodied perspective extends beyond viewing emotion as a cognitive appraisal of external stimuli, positioning it instead within dynamic, reciprocal loops between sensory input, interoceptive states, and motor expression (Niedenthal, 2007; Barrett, 2017). Affective images, whether photographic depictions, artistic renderings, or controlled laboratory scenes, serve as potent elicitors that engage these loops, influencing facial muscle activation, posture, heart rate variability, and respiratory patterns, which in turn modulate the appraisal and intensity of emotional responses (Critchley & Garfinkel, 2017).

Within experimental psychology, affective image-based paradigms are crucial both for standardisation and ecological validity. The International Affective Picture System offers reproducible stimuli with normative ratings, yet novel approaches have introduced immersive visual environments, dynamic stimuli, and culturally relevant imagery to elicit emotions that resonate more deeply with embodied states (Lang et al., 2008; Mikels et al., 2005). These developments allow researchers to examine how sensorimotor resonance, proprioceptive feedback, and interoceptive accuracy contribute to the generation and regulation of emotion (Craig, 2009). Such methodologies not only clarify the mechanisms underpinning emotional experience but also illuminate individual and cultural differences in bodily involvement during affective processing.

Practical applications follow naturally from these insights. In clinical contexts, guided imagery incorporating bodily engagement, through posture adjustment, breath patterns, or weighted tactile feedback, has shown promise in modulating affective states for individuals with anxiety and mood disorders (Gross, 2015; Matsumoto et al., 2006). In resilience training and wellbeing interventions, carefully designed affective images can evoke adaptive bodily states that strengthen positive affect and reduce stress reactivity. These approaches bridge laboratory research and applied psychology, demonstrating that the systematic use of affective imagery in embodied contexts offers a scientifically grounded, versatile tool for both understanding and enhancing human emotional life.

# 6.2 Visual Stimuli and Embodied Emotions: Foundations and Importance

Visual stimuli have long played a pivotal role in emotion research, offering insights into the nuanced interplay between perception, cognition, and emotion. Historical investigations have revealed how visual imagery elicits diverse emotional reactions, demonstrating that images are potent tools for emotional engagement and research. Early psychological experiments by pioneers such as Ekman (1973) and Friesen (1972) underscored the universality of facial expressions, suggesting that basic emotions could be reliably elicited and recognised across cultures through visual means. These foundational studies established visual stimuli, especially facial imagery, as critical tools for investigating emotional processes.

The theoretical basis for using visual stimuli in emotion research emerges from the understanding of embodied cognition, where emotional states are not just psychological phenomena but are deeply connected to physical bodily states and environmental interactions. According to this framework, visual stimuli do not merely evoke emotional states. They actively shape bodily responses, physiological reactions, and subsequent emotional experiences. Figure 1 summarises this flow, emphasising the embodied link between what is seen and how the body reacts. For instance, viewing a fearful face can trigger autonomic nervous system responses such as increased heart rate or skin conductance, reflecting embodied emotional activation.

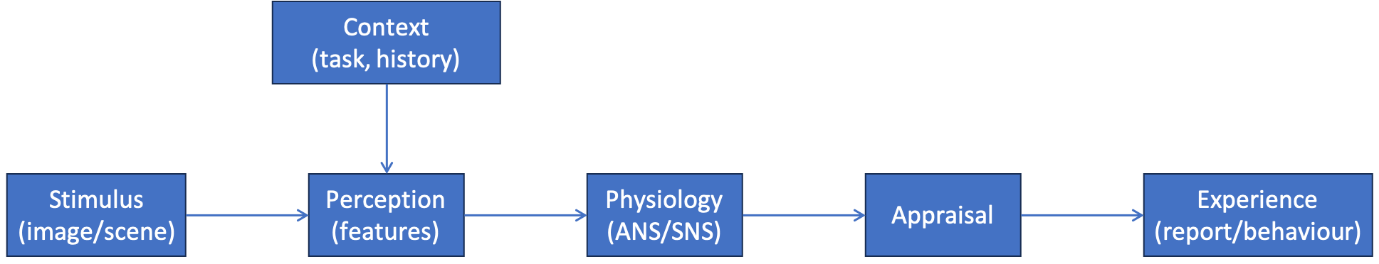


Figure 1. From visual stimulus to embodied emotion

Contemporary research extends this theoretical framework by exploring complex interactions between visual cues and bodily expressions, highlighting how context modifies emotional perception. Studies involving dynamic interactions between visual stimuli (e.g., emotional scenes, facial expressions, and body postures) emphasise the importance of ecological validity in eliciting genuine emotional responses. These investigations reveal that emotional experiences are not static but contextually driven, dynamically influenced by the integration of facial, bodily, and environmental cues.

The significance of visual stimuli in eliciting embodied emotions lies in their ability to replicate naturalistic emotional contexts within controlled laboratory settings. This capability is invaluable for experimental rigour, providing researchers with precise control over stimulus parameters while capturing authentic emotional reactions. Consequently, visual stimuli have become indispensable in advancing our understanding of embodied emotional processes, forming the cornerstone of contemporary emotion research and practical applications in fields such as psychotherapy, affective computing, and media psychology.

# 6.3 Embodied Emotion Elicitation: Principles and Techniques

Emotion elicitation lies at the core of embodied emotion research, probing ways to provoke real, lived emotional reactions in the body, not merely in subjective self-report. Two key methods have emerged as complementary yet distinct: **priming** (a subtle preparatory exposure that implicitly influences subsequent emotional responses) and **provoking** (the deliberate presentation of emotionally charged stimuli to trigger robust bodily and psychological reactions). Figure 2A sketches a brief priming sequence, while Figure 2B shows a sustained provoking sequence.

|  |
| --- |
|  |
|  |

Figure 2. Families of elicitation designs: (A) Priming timeline; (B) Provoking timeline

**Priming** involves exposing individuals to brief cues (often outside of awareness) that subtly bias emotional responses later. Affective priming research illustrates how valenced stimuli (words, images, or odours) rapidly influence subsequent behaviour or perception even when not consciously perceived (Rohr et al., 2021). For example, masked emotional words can accelerate the recognition of congruent targets or reduce cognitive interference (Visalli et al., 2023). In contrast, **provoking** consists of direct exposure to emotionally intense stimuli, such as film clips, evocative scenes, or immersive virtual environments. Film-based induction, standardised image sets (International Affective Picture System, IAPS), and Virtual Reality (VR) scenarios reliably evoke strong emotional responses measurable via physiological and behavioural channels (Boğa et al., 2023). Each method serves different experimental goals: priming is ideal when subtle shifts or non-conscious modulation are crucial; provoking suits investigations of full-bodied emotional states, high arousal responses, or ecological validity.

Embodied emotion theories propose that emotions are grounded in bodily states, sensorimotor simulations, and interoception (Barsalou, 2008; Reddan et al., 2024). The **embodied simulation framework** asserts that perceiving emotion triggers partial reenactments in somatosensory and motor cortices, providing a physiological pathway for emotion elicitation via bodily resonance (Reddan et al., 2024). Behavioural mimicry, such as facial muscle activation during perception, has been shown to facilitate emotional experience (Niedenthal, 2007).

**Interoceptive priming** also plays a key role since bodily signals like heartbeat, respiration, and visceral awareness modulate emotional readiness. Salamone et al. (2021) demonstrated that priming interoceptive sensations engages insular and anterior cingulate networks, influencing subsequent emotional interpretation (e.g., rating of images).

From a psychophysiological perspective, embodied elicitation relies on autonomic and somatic signatures: heart rate variability (HRV), electrodermal activity (EDA), facial electromyography (EMG), and pupillometry may all change in response to even implicit priming. As recognised in modern affective computing, these signals map bodily emotional responses central to both priming and provoking techniques.

Emotion elicitation methods are not merely academic. They underpin applications across multiple domains.

1. **Affective computing and artificial intelligence (AI) system training.** Emotion algorithms require rich, embodied emotional data. Using priming and provoking methods to elicit genuine emotion helps in building datasets like DECEiVeR (Aly et al., 2024) and VR-based platforms such as the Magic Xroom (Mousavi et al., 2023), where stimuli adapt in real time to evoke flow-related emotional states. This trajectory aligns with the ‘Artificial Psychology’ (PsAIchology) framing that integrates AI methods within psychological science (Farahani et al., 2024).
2. **Clinical and therapeutic interventions.** Subtle priming techniques, such as odor-word priming or memory-based positivity primes, can modulate mood or reduce negative emotional biases, offering a gentle, potentially therapeutic route (Cereghetti et al., 2024; Bargh et al., 1996).
3. **Education and simulation training.** In simulation education, recognising emotional states via physiological feedback (heart rate, facial muscle activation) and elicitation through scenario priming enhances learners’ emotional engagement and learning outcomes (LeBlanc et al., 2024).
4. **Cognitive and social psychology research.** Emotional priming influences cognitive control (e.g., emotional Stroop paradigms), decision-making, and social behaviour (Visalli et al., 2023; Dennison, 2024).

To clarify practical considerations, Table 1 shows a high-level comparison highlighting their respective strengths:

|  |  |  |
| --- | --- | --- |
| Aspect | Priming | Provoking |
| Awareness | Often subliminal or implicit | Explicit and consciously perceived |
| Emotional Intensity | Subtle, low-arousal modulation | Strong, high-arousal emotions |
| Physiological Signature | Minimal but measurable (e.g., EDA fluctuations) | Robust autonomic response (HR, facial EMG, pupillometry) |
| Application Contexts | Mood modulation, non-conscious bias, social priming | Film stimuli, VR scenarios, IAPS images |
| Temporal Effects | Small shifts in baseline emotional readiness | Immediate emotional experience |

Table 1. Comparative characteristics of priming and provoking methods for emotion elicitation

Priming is especially powerful in studies of personality or clinical styles: for example, depressed individuals show greater bias following negative primes and slower adjustment to positive primes (Rohr et al., 2021). Provoking, in contrast, enables immersive study of full-bodied emotional episodes using film, VR, or live scenarios (Boğa et al., 2023; Polo et al., 2025).

Eliciting emotion should always occur within an ethically guided framework. Priming often escapes conscious awareness, raising questions of consent. Thus, participants must be fully informed indirectly, without revealing prime content. Provoking may evoke intense psychological states; thus, debriefing, emotional support, and screening protocols are essential. Because emotion elicitation impacts physiological systems, measurement must be conducted sensitively and respectfully. Robust protocols for data privacy, participant autonomy, and emotional risk minimisation are fundamental.

Embodied emotion elicitation rests on a principled integration of priming and provoking methods, grounded in physiology, somatosensory simulation, and interoceptive feedback. These methodologies enable researchers to evoke and study authentic emotional states across multiple contexts, from subtle mood shifts to immersive experiences, informing fields from affective computing to clinical intervention and education. With appropriate ethical safeguards and multimodal physiological measurement, emotion elicitation continues to drive forward our understanding of how emotional experiences are embodied, enacted, and measured in profound ways.

# 6.4 Facial Expressions, Body Language, and Contextual Influences on Emotional Experience

The perception and elicitation of emotion are deeply embodied processes, with facial expressions, body language, and contextual cues operating as essential components of the emotional communication system. Visual cues, especially those derived from the human face and posture, serve as crucial mediators in the embodied experience of emotion, influencing not only how emotions are conveyed but also how they are internally experienced by both the emitter and the observer. In this section, we examine the integrative role of these nonverbal modalities, highlighting the dynamic interplay between facial and bodily signals and the surrounding context in shaping emotional perception and elicitation.

Facial expressions are widely recognised as primary channels for nonverbal emotional communication. Rooted in Darwinian theories of evolution, facial expressions are often viewed as biologically hardwired signals designed to convey survival-relevant information (Darwin, 1872; Ekman, 1992). Classic research has long emphasised the universality of six basic emotions (happiness, sadness, anger, fear, surprise, and disgust), each associated with specific facial muscle configurations. However, recent evidence complicates this narrative by highlighting the variability and contextual dependency of these expressions.

Jack et al. (2012) used dynamic three-dimensional modelling and machine learning approaches to reveal that cross-cultural perception of facial emotion differs substantially, challenging the universality hypothesis. Research using deep learning facial recognition technologies has shown that micro-expressions (transient, involuntary facial movements) can offer rich insight into underlying affective states that may not align with overt emotional expressions (Yan et al., 2013). In the context of embodied emotion research, facial mimicry also emerges as a key mechanism. When individuals unconsciously imitate another’s facial expression, it fosters affective resonance and emotional alignment. This is a process crucial for empathy and social bonding (Wood et al., 2016). This mimetic mirroring facilitates embodied simulation, a neural mechanism through which observers internalise the emotional states of others (Gallese & Sinigaglia, 2011).

Beyond the face, the body provides an expansive canvas for emotional signalling. Posture, gesture, and kinetic energy in movement all contribute to how emotions are perceived and felt. For example, open and expansive postures often indicate confidence or joy, while contracted or defensive postures can signal fear, shame, or anxiety (Carney, Cuddy, & Yap, 2010; de Gelder, 2016). Even in the absence of facial information, observers can often accurately discern emotional states based solely on body configuration and motion (Atkinson et al., 2004; de Gelder et al., 2015; Blythe et al., 2023).

Recent embodied cognition models suggest that body posture not only reflects emotional states but also actively influences them. In one study, participants asked to adopt certain postures experienced congruent shifts in mood and self-appraisal, a phenomenon known as postural feedback (Stepper & Strack, 1993; Elkjær et al., 2022). While some recent replications have questioned the magnitude of these effects (Cesario, Jonas, & Carney, 2017), evidence continues to support the notion that posture contributes to emotional embodiment, particularly in high-stakes or socially evaluative situations.

Body expressions are crucial in situations of emotional ambiguity, where facial expressions alone are insufficient. For instance, de Gelder (2016) argues that whole-body expressions act as holistic emotional units that often override or reinterpret facial cues. In virtual agents, combining facial + vocal/body cues can shape trust and recognition (Torre et al., 2019; Tsiourti et al., 2019). While facial and bodily signals are vital, their interpretation is rarely isolated from contextual information. Context modulates the perception of emotional cues, refining or even inverting the affective message. In naturalistic settings, emotional expressions are interpreted within a complex web of social, cultural, spatial, and temporal variables. For example, Aviezer et al. (2012) demonstrated that observers often rely more on body posture than facial expression when faced with incongruent emotional cues (e.g., a smiling face on a defeated body). These findings suggest that bodily context not only colours emotional interpretation but can dominate the emotional appraisal process.

Cultural and situational norms also play a crucial role. Park and Han (2018) found that collectivist cultures tend to emphasise contextual cues over facial expressions when decoding emotions, whereas individualistic cultures prioritise facial cues. This has implications for designing AI-based emotion recognition systems in multicultural environments, where facial data alone may yield biased or misleading interpretations.

Temporal dynamics, including the timing and sequencing of expressions, also influence how emotions are perceived. Dynamic facial expressions, which unfold over time, tend to be more accurately recognised than static ones (Krumhuber et al., 2021). Similarly, the preceding emotional context of an interaction shapes anticipatory biases, where observers expect congruent emotional expressions and may misinterpret ambiguous signals to fit those expectations (Barrett, 2017).

Affective understanding is fundamentally multimodal. Rather than operating as discrete channels, facial expressions, body language, and contextual features are integrated into a cohesive perceptual experience. The embodied approach posits that this integration occurs not merely at a cognitive level, but within the perceptual-motor and visceral systems of the observer (Niedenthal et al., 2005). The perceiver's body plays a constitutive role in emotion perception, with sensorimotor resonance (i.e., bodily simulation of another’s posture or expression) guiding emotional inferences.

Neuroimaging implicates interoceptive and premotor circuits during bodily emotion perception (Azzalini et al., 2019), consistent with predictive-processing accounts (Clark, 2013). These findings align with predictive processing models, which suggest that the brain uses embodied priors to anticipate and interpret emotional signals in a top-down manner (Clark, 2013).

# 6.5 Methods of Embodied Emotion Priming and Provoking

The scientific exploration of emotion necessitates methodologies that can reliably and ethically elicit affective responses in controlled environments. Among these, **priming** and **provoking** methods represent two foundational strategies that serve distinct psychological and neurobiological purposes in embodied emotion research. While both aim to evoke emotional states, they differ significantly in subtlety, intensity, and temporal dynamics. This section outlines the theoretical underpinnings, procedural approaches, and ethical considerations for both methods, emphasising their application in embodied emotional studies and affective computing.

Priming refers to the psychological technique of preparing an individual’s emotional or cognitive system by exposing them to specific, often subtle cues that influence their subsequent affective experience or behaviour without conscious awareness. In the context of embodied emotion, priming is particularly useful for engaging pre-existing emotional schemas without overt stimulation, allowing researchers to explore how emotions arise from implicit cognitive-affective networks. Several paradigms have emerged to prime emotional states:

* **Semantic priming**, wherein participants are exposed to emotionally-valenced words (e.g., death, joy) that bias their responses to subsequent stimuli (Elgendi et al., 2018);
* **Affective visual priming**, where briefly presented images (such as facial expressions, affective scenes, or symbolic cues) influence perception or judgment, even if presented subliminally (Elgendi et al. 2018; Nava & Turati, 2020);
* **Embodied priming**, which leverages bodily manipulations, such as facial muscle activation or posture, to bias affective experience and related physiology (Price & Harmon-Jones, 2015; Van Cappellen, Ladd, Cassidy, Edwards, & Fredrickson, 2022).

Recent findings suggest that priming not only influences affective judgments but can modulate physiological responses such as heart rate variability (Carlisle et al., 2012; Li et al., 2007) and skin conductance (Nava & Turati, 2020). These embodied correlates reveal that priming is not merely a cognitive effect but engages the broader nervous system, aligning with the principles of embodied cognition.

In contrast to priming, provoking methods are designed to **explicitly and robustly evoke emotional states**, often through multimedia stimuli that are rich in affective content. These methods are used when stronger, more observable emotional reactions are required, particularly in studies involving psychophysiological measurement, neuroimaging, or behavioural change. The most commonly used tools include:

* **Standardised image databases** such as IAPS and OASIS (Kurdi et al., 2017), which provide a range of affective scenes normed for valence and arousal;
* **Emotion-inducing film clips**, such as those from the Database for Emotional Movie Stimuli or the FilmStim (Schaefer et al., 2010), deliver continuous, multimodal emotional experiences;
* **Auditory stimuli**, including affective music and sound effects, have proven effective in eliciting sadness, fear, and joy, often used in immersive settings (Khalfa et al., 2005);
* **VR-based emotional scenarios**, which provide high ecological validity and embodiment, enabling participants to engage emotionally through real-time, immersive interaction (Somarathna et al., 2022).

In embodied research, provoking methods are increasingly paired with **real-time physiological monitoring**, such as electroencephalography (EEG), EMG, heart rate variability, and galvanic skin response, to detect the embodied correlates of the elicited emotions (Shu et al., 2018). These measurable responses allow for objective validation of emotional states, enriching our understanding of how external stimuli translate into embodied emotional experiences.

Both priming and provoking techniques operate on shared but differently emphasised **cognitive-affective and sensorimotor pathways**.

* **Priming** functions primarily via **top-down** cognitive processes. It activates associative memory networks and predictive coding mechanisms that prepare the individual for an expected emotional context (Barrett, 2017). This anticipatory activation often remains below the threshold of awareness, influencing perception and interpretation;
* **Provoking**, by contrast, engages **bottom-up sensory integration** mechanisms. Emotional stimuli directly trigger autonomic and limbic responses, particularly involving the amygdala, insula, and brainstem nuclei, that manifest as observable physiological and affective reactions (Pessoa, 2013).

Importantly, the **embodiment hypothesis** suggests that both priming and provoking activate internal simulations of emotional experience via **sensorimotor re-enactment**. This aligns with theories of grounded cognition, wherein emotions are not only mental states but bodily states reactivated in perception and action (Niedenthal, 2007; Baumgartner et al., 2006).

Given the **potential vulnerability** of participants during emotion elicitation studies, ethical protocols must be rigorously followed. The following considerations are paramount:

1. **Informed consent** must explicitly state the possibility of emotional discomfort and detail the nature of the stimuli, especially when using provocative methods;
2. **Debriefing procedures** should be comprehensive, offering emotional support and the option to withdraw data if participants feel distressed;
3. **Use of trauma-sensitive design** is crucial when employing emotionally intense stimuli. Participants with a history of trauma or psychological conditions should be pre-screened or offered opt-outs;
4. **Minimising deception** is essential, particularly for priming techniques that may operate subliminally. While some degree of masking may be needed for methodological validity, it should never compromise participant safety;
5. **Institutional oversight**, such as approval from ethics committees or Institutional Review Boards, is mandatory for all studies involving emotional elicitation.

Recent guidelines from affective computing and human-computer interaction research emphasise **participant agency, emotional aftercare, and transparency** (Mohammad, 2022) in the use of AI-driven affective systems. For example, studies involving virtual agents designed to provoke emotional responses must be especially cautious to avoid manipulative interaction patterns or emotionally misleading cues. To ensure both scientific validity and ethical soundness, the following best practices are recommended:

* Use **multi-method validation**, combining self-report, behavioural, and physiological measures;
* **Calibrate stimuli** using normative databases to match participant demographics and cultural backgrounds;
* Employ **randomised controlled designs** to reduce bias and habituation.
* Pilot test all priming and provoking materials in smaller samples to assess tolerability and effect size;
* Incorporate **adaptive algorithms** in real-time systems that monitor participant distress and adjust stimuli intensity accordingly.

Figure 3 shows a calibration scatter in the valence–arousal space, which we use to verify balanced coverage and avoid clustering stimuli in one quadrant before running the study.

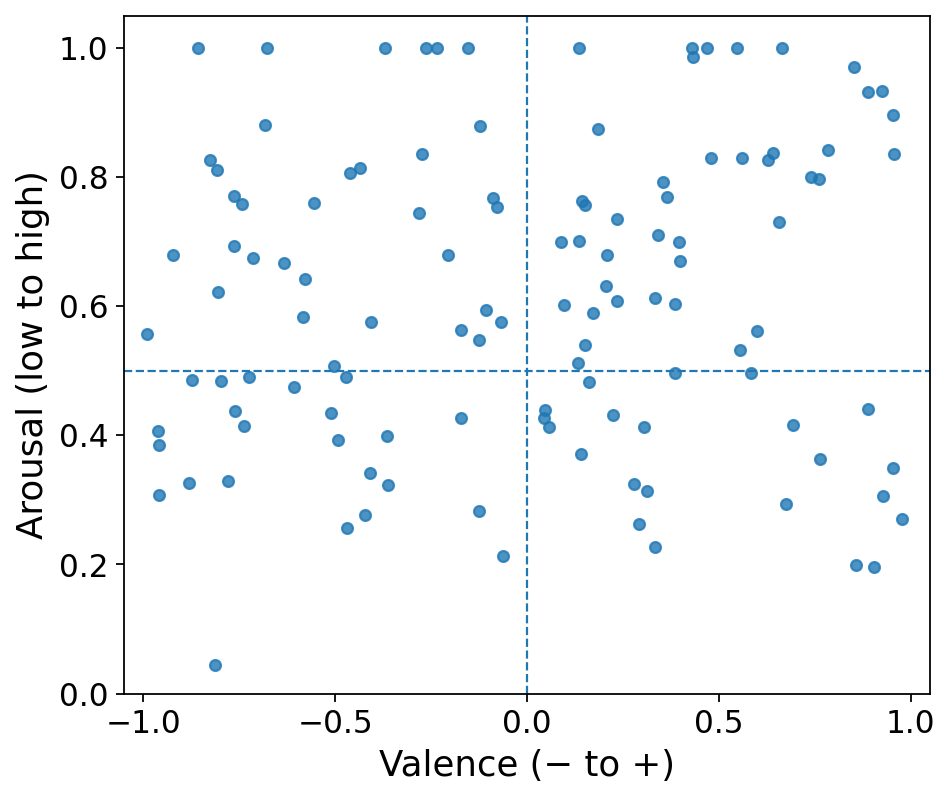


Figure 3. Stimulus calibration map (valence × arousal)

Figure 4 sketches the adaptive loop we use in practice: continuous measurement (e.g., EDA/HR) feeds a decision policy that either pauses or adjusts intensity, with clear stop criteria.

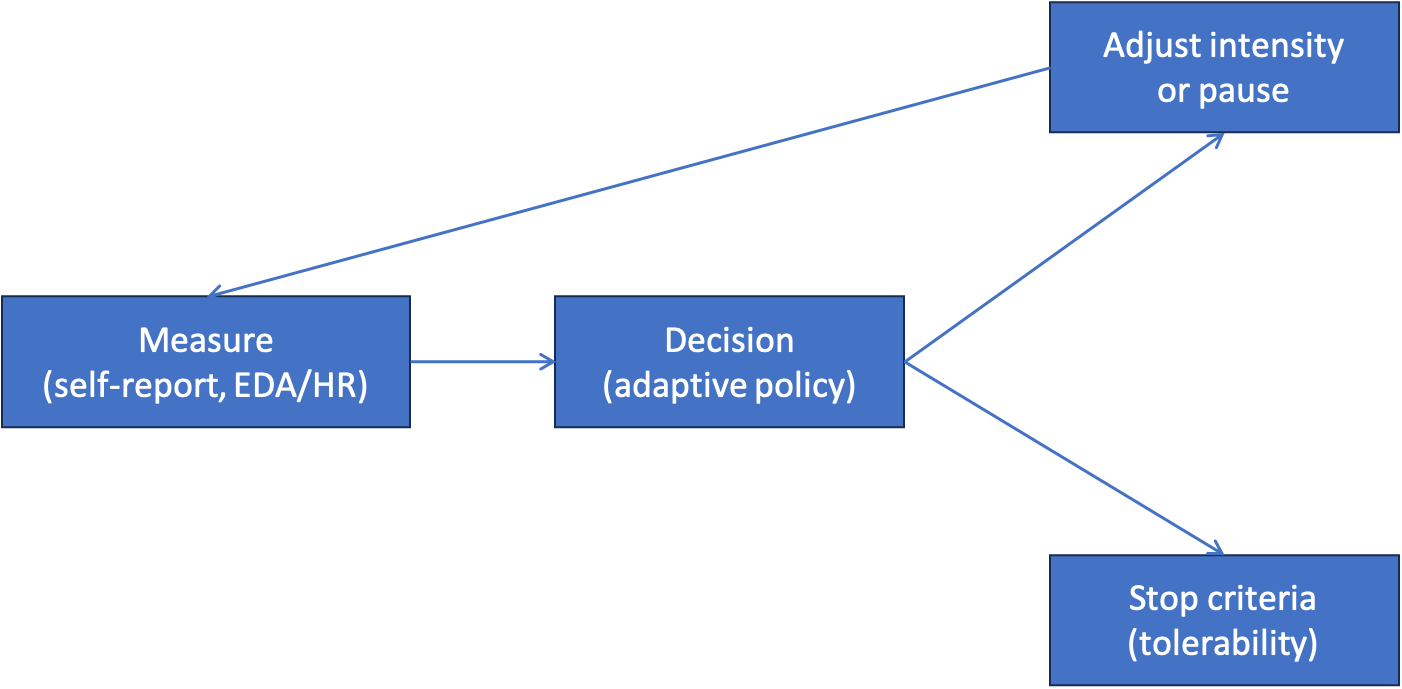


Figure 4. Adaptive elicitation loop (monitor → decide → adjust)

**Priming and provoking** are complementary methods of emotional elicitation that, when used ethically and precisely, can reveal the complex interplay between cognition, emotion, and embodiment. Their integration into affective science and applied technologies, such as therapeutic interfaces and emotionally aware AI, marks a significant step toward more nuanced and responsive emotional systems. As affective research becomes increasingly multimodal and personalised, these methods will remain foundational tools in our quest to understand and model the embodied human emotional experience.

# 6.6 Computational Analysis of Affective Images and Emotional Elicitation

In the last decade, the field of emotion research has been transformed by advances in computational methods, especially those powered by artificial intelligence and machine learning. The capacity to automatically analyse and interpret affective visual stimuli has not only expanded the scale and efficiency of affective research but also deepened our understanding of embodied emotion. Rather than relying solely on subjective self-report or manual coding, researchers can now draw on sophisticated algorithms to extract and classify emotional information from facial expressions, bodily postures, and even contextual visual cues. These methods are particularly vital in complex experimental setups where continuous emotional monitoring is needed or where real-time emotion modelling is essential, such as in virtual agents, adaptive learning environments, or therapeutic interventions. For psychology-focused, Python-based modelling patterns and regression pipelines, see (Kovač et al., 2024).

At the core of this computational turn lies the development of emotion recognition systems that leverage deep learning architectures, particularly convolutional neural networks (CNNs), to identify subtle affective patterns within static images or video streams. Facial expression recognition (FER) has emerged as one of the most widely explored applications in this area. Using large annotated datasets like AffectNet (Mollahosseini et al., 2017), RAF-DB (Li et al., 2017), and FER2013, researchers train models that can predict basic and compound emotions based on the activation of facial landmarks, muscle dynamics, and temporal transitions.

The move toward real-time FER systems has further improved ecological validity in emotion elicitation studies, allowing researchers to monitor facial affect as participants engage with emotionally provocative stimuli. To make the processing steps concrete, Figure 5 shows a compact FER pipeline (preprocess → CNN encoder → classifier) alongside an illustrative saliency heatmap highlighting face regions that drive a predicted label.

|  |  |
| --- | --- |
|  |  |

Figure 5. Facial emotion recognition overview: (A) FER pipeline; (B) saliency heatmap indicating image regions most influential for the prediction

Yet emotion is rarely confined to the face. Recent models increasingly emphasise the importance of full-body information in emotion understanding. These multimodal systems integrate facial features with body language cues, such as arm gestures, posture openness, and motion fluidity, to more accurately decode affective states. Figure 6 connects this multimodal view to our conceptual model by showing how face, body, and scene jointly inform perceived emotion.

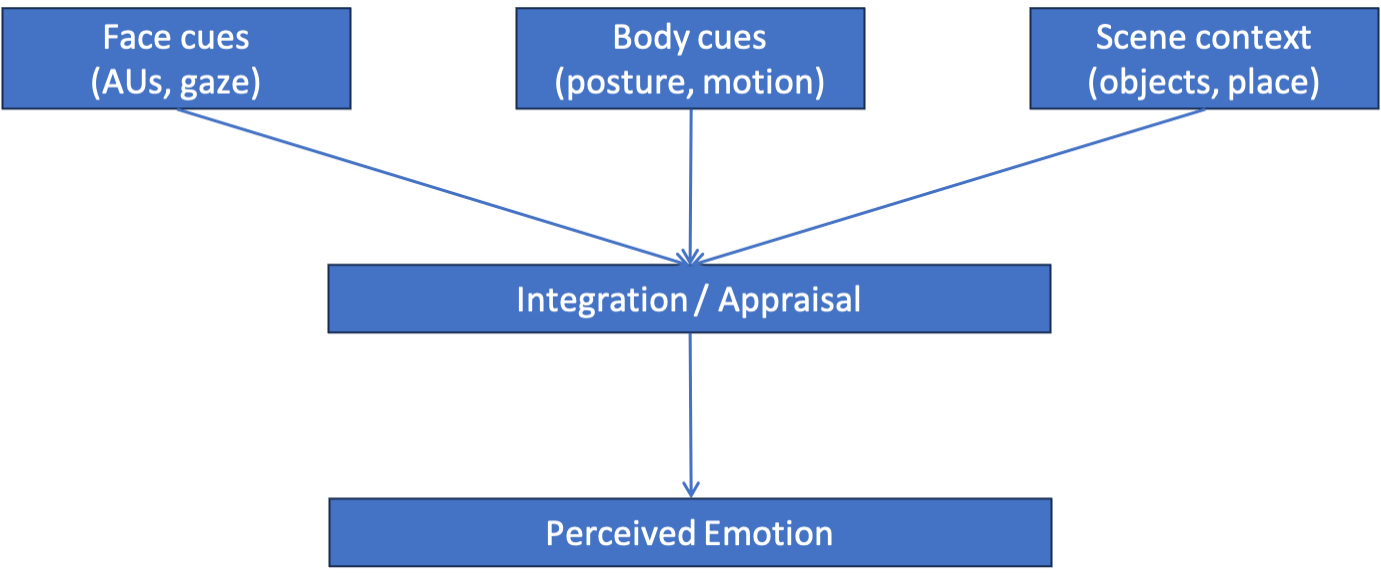


Figure 6. Context matters: face + body + scene integration

Temporal modelling techniques like Long Short-Term Memory networks and Transformer-based architectures allow these systems to maintain a memory of gesture progression, enabling better interpretation of dynamic emotional behaviour over time. This is particularly useful in contexts where the emotional response is not immediate but unfolds across several moments, such as during narrative-based video stimuli or in therapy simulations using avatars (Kosti, Álvarez, Recasens, & Lapedriza, 2019; Mittal et al., 2020).

A crucial innovation in this space involves context-aware emotional modelling. While early systems assumed that emotions are universal and static, current trends recognise the layered nature of emotional experience, which is often shaped by surrounding environmental cues. For example, a smiling expression in a funeral scene may carry a completely different emotional meaning than the same expression in a birthday party. To capture these differences, AI models now incorporate contextual embeddings (extracting background, object co-occurrence, and temporal consistency) to modify predictions based on visual surroundings. To capture these differences, current models incorporate contextual cues, such as background, objects, and global scene, to refine predictions. Multi-branch CNNs and context-aware networks that fuse person and scene information consistently outperform person-only baselines (Kosti et al., 2019; Lee, Kwon, & Plataniotis, 2019). Related machine-learning frameworks have also been used for affect-linked constructs beyond core emotions, such as love addiction, where feature patterns and explanations clarify predictive factors (Farahani et al., 2025). Building on this, Figure 7 contrasts early fusion (concatenate features, then classify) with late fusion (separate classifiers whose scores are fused), a practical choice when combining face, body, and scene/context streams.

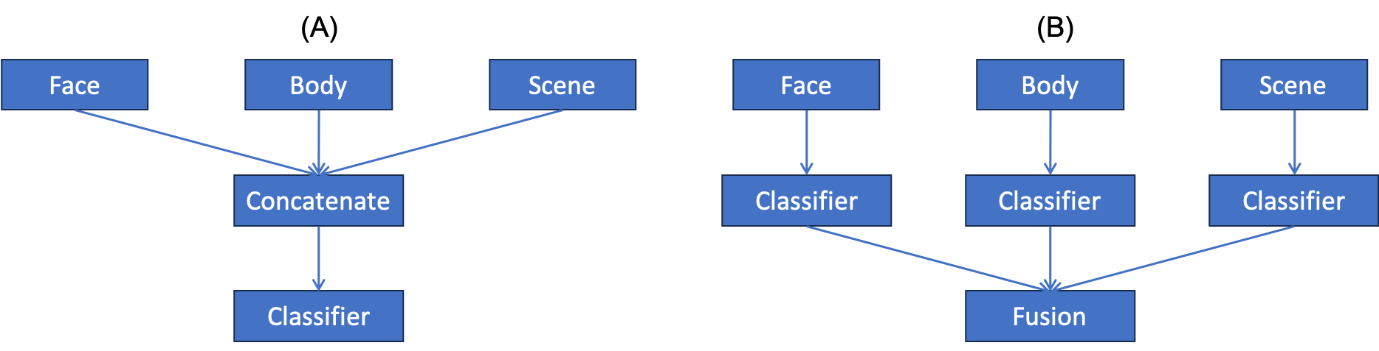


Figure 7. Strategies for combining face, body, and scene features in multimodal emotion recognition: (A) early-fusion; (B) late-fusion

Beyond recognition, computational techniques are also revolutionising how emotion is elicited. Generative AI tools like Generative Adversarial Networks and diffusion models are used to synthetically generate affective images tailored to specific emotional categories. These generated stimuli offer several advantages: they are controllable, ethically safe, and capable of spanning the full valence-arousal space with fine-grained precision. For instance, rather than sourcing distressing real-world images to provoke anxiety, researchers can create synthetic scenes that are sufficiently evocative without being ethically problematic. Emotion-generative agents are being tested in VR and Augmented Reality (AR) settings, where AI-driven virtual characters engage users in emotionally rich conversations, adjusting their expressions and gestures based on user feedback (Galanos et al., 2021).

The use of computational methods for emotional elicitation demands rigorous validation. While model accuracy remains a primary concern, embodiment researchers also investigate the degree to which AI-predicted emotions correspond with physiological signals, such as heart rate variability, pupil dilation, or galvanic skin response. In these studies, emotional predictions are considered valid not only when they match participant self-reports but also when they synchronise with autonomic nervous system responses. This dual benchmarking ensures that computational analyses are grounded in embodied emotional realities, not just visual classification heuristics. A number of case studies illustrate these methods in practice. Comparable supervised pipelines are also applied to psychological health outcomes, for example the classification of chronic pain (Kovač et al., 2025). A complementary line of work uses calibrated regression to predict internal shame, with XGBoost performing best and distress tolerance emerging as the strongest predictor (Kovač, Ratković, Farahani, & Watson, 2025b). For example, end-to-end multimodal architectures combining facial video with other streams (e.g., audio, text, or gaze) improve recognition in naturalistic settings compared with facial-only models (Tzirakis, Trigeorgis, Nicolaou, Schuller, & Zafeiriou, 2021; Gu, Zhang, & Ma, 2018). In another project, real-time EEG classification was used to monitor affect during video stimuli and could be integrated to adapt stimulus delivery on the fly (Nandi, Saha, & Acharjee, 2021).

Such examples demonstrate that computational systems are not only observers of affect but can act as **interactive co-regulators** of emotional experience. This aligns closely with embodied theories of emotion, where affect is not a static label but a process dynamically co-constructed between agent and environment. The future of emotional elicitation lies in these intelligent, responsive systems that adapt to the user’s embodied state in real time.

Yet challenges remain. Bias in training datasets, particularly in facial expression data across cultures and skin tones, continues to undermine the generalizability of many models. Ethical debates are also mounting regarding the manipulation of emotional states via automated systems, particularly in commercial applications where user consent and emotional safety may be compromised. To navigate this, emerging frameworks propose transparency, participant control, and real-time opt-out mechanisms in any system capable of emotional modulation.

As AI continues to evolve, the integration of computational methods into embodied emotion research opens new frontiers for inquiry and application. From designing personalised therapeutic interventions to enhancing the affective responsiveness of artificial agents, these tools offer powerful means of decoding and modulating human emotion. The key, however, lies in grounding these technologies in robust psychological theory and embodied ethics, ensuring that the machines we build recognise emotion not just as data, but as human experience.

6.7 Measuring and Quantifying Embodied Emotional Responses

Understanding embodied emotion requires not just the elicitation and recognition of emotional states, but also their rigorous measurement and quantification. Emotion, inherently multifaceted, manifests across multiple physiological, behavioural, and computational domains. While subjective self-report remains a valuable source of emotional insight, it is insufficient when investigating real-time, unconscious, or embodied affective processes. Therefore, contemporary emotion research employs a unified methodological framework that integrates physiological sensors, behavioural observations, and algorithmic quantification tools to provide a multilayered and objective understanding of emotional experience. This section synthesises core methods for measuring embodied emotion and discusses their complementarities, applications, and challenges in both lab-based and naturalistic settings.

Among the most established techniques in the physiological domain is HRV, a widely accepted index of autonomic nervous system function. HRV captures the variation in time intervals between successive heartbeats, offering insight into sympathetic and parasympathetic activity. Emotional states characterised by arousal, such as fear or excitement, often lead to reductions in HRV, while relaxed or positive affective states are associated with greater variability (Shaffer & Ginsberg, 2017). Tools like the Polar H10 and Empatica E4 wristband allow for non-invasive, continuous HRV monitoring (Umair et al., 2021), making them ideal for both lab-based emotion induction tasks and field studies. HRV is particularly well-suited for embodied research due to its close ties to vagal tone and the neurovisceral integration model, which positions bodily regulation at the centre of affective processing (Smith et al., 2017).

EEG provides a more direct window into the brain's electrical activity during emotional processing. It is especially useful for examining the temporal dynamics of emotion, capturing millisecond-level changes in cortical activation. Specific EEG markers, such as frontal alpha asymmetry, have been consistently associated with approach or avoidance motivational tendencies, thereby linking emotion to embodied action schemas (Coan & Allen, 2004; Harmon-Jones et al., 2010). In more advanced setups, source localisation techniques like sLORETA can be used to estimate the origin of emotional signals within cortical networks, enhancing the spatial interpretability of EEG data. Mobile EEG systems are increasingly deployed in ecologically valid settings, such as immersive VR or real-world social interactions, enabling the study of embodied emotion beyond the confines of the lab.

Another powerful physiological method is **skin conductance response (SCR)**, often used to measure sympathetic nervous system arousal. SCR detects changes in sweat gland activity driven by emotional arousal. Unlike HRV or EEG, which may reflect both valence and arousal, SCR is a relatively pure index of arousal intensity, regardless of whether the emotion is positive or negative. This makes it an excellent tool for measuring emotional responses to provocative stimuli, such as fear-inducing images or emotionally charged film scenes. Importantly, combining SCR with facial and behavioural measures enhances interpretative power, as high arousal alone does not distinguish between anger, fear, or excitement.

Beyond physiological signals, **behavioural and observational measures** provide critical insights into how emotions are embodied and expressed. **Eye-tracking**, for example, allows researchers to examine visual attention patterns and pupil dilation during emotional tasks. Participants tend to fixate more quickly and longer on emotionally salient regions, such as the eyes or mouth in faces, particularly when processing threat or socially relevant cues (Vetter, Badde, Phelps, & Carrasco, 2019; Capriola-Hall, Ollendick, & White, 2021). Changes in **pupil size** have also been linked to affective arousal and cognitive effort, providing a dual index of embodied emotional engagement (Mathôt, 2018; van der Wel & van Steenbergen, 2018).

Perhaps one of the most influential tools in behavioural emotion research is the **Facial Action Coding System (FACS)**, originally developed by Ekman and Friesen (1978) and now often implemented with automated software such as OpenFace. FACS identifies distinct facial muscle movements (action units) and maps them onto specific emotional expressions. For example, the combination of AU6 (cheek raiser) and AU12 (lip corner puller) is strongly associated with genuine (Duchenne) smiling. Automated FACS systems offer a non-invasive and real-time means of capturing facial expressivity, which is particularly relevant in studies of emotion contagion, mimicry, or emotion regulation.

**Postural and gesture analysis** is another growing area of embodied emotion measurement. Advances in depth sensing and motion capture technologies, such as Microsoft Kinect (Zhang, 2012) and OpenPose (Kim et al., 2021), allow for real-time tracking of body configuration, movement fluidity, and gesture dynamics. These kinematic features can then be analysed to infer emotional states. For instance, slow, constricted movements may indicate sadness or fear, while expansive, energetic gestures are often linked to happiness or anger. Integrating posture data with facial and vocal cues creates a richer and more holistic model of embodied emotion.

On the computational front, emotion quantification has been significantly advanced by the use of **multimodal machine learning models**. These systems integrate data from physiological signals, facial expressions, and behavioural features to predict emotional states with high accuracy. Ensemble methods, recurrent neural networks, and attention-based architectures like Transformers are now employed to fuse temporal data streams and classify emotions dynamically. Emotion recognition challenges such as DEAP (Koelstra et al., 2011) and AMIGOS (Miranda-Correa et al., 2018) have provided benchmark datasets that pair EEG, SCR, HRV, and video data with self-reported emotional labels, enabling researchers to train models that learn the temporal and cross-modal dependencies inherent to emotional responses.

Quantifying embodied emotion also involves estimating **valence-arousal dimensions** rather than discrete emotion labels. Models like the Circumplex Model of Affect (Russell, 1980) conceptualise emotions along continuous dimensions of pleasure and activation, which can be mapped using regression-based models on physiological and behavioural inputs. For example, a system may output a high arousal–low valence profile, suggesting a state of anxiety or anger. These dimensional outputs are especially useful in adaptive systems that need to respond to gradients of emotion rather than fixed categories.

Despite these advancements, challenges remain in aligning computational predictions with subjective experience. Embodied emotional responses are highly individual, shaped by personal history, cultural background, and context. Personalised calibration of emotion recognition systems, through baseline recordings, user-specific models, or adaptive learning, is crucial for achieving ecological validity and ethical sensitivity. As emotion measurement becomes more automated and multimodal, integrating these methods offers a more comprehensive and embodied view of affect. Physiological data captures internal states, behavioural observations reflect external expressions, and computational modelling provides the analytical link that connects these layers. Collectively, they enable a rich, multidimensional understanding of how emotion is not just felt or reported, but experienced through the body, expressed in movement, and interpreted by intelligent systems.

6.8 Embodied Emotion Elicitation in Digital and Virtual Environments

As digital interfaces become more immersive, the elicitation of emotion in virtual contexts has shifted from a passive process to one that is fundamentally embodied and interactive. VR, AR, and emotionally intelligent digital agents now offer unprecedented opportunities to simulate, modulate, and analyse emotional experiences with a level of ecological validity previously unattainable in traditional laboratory settings. These technologies not only improve the realism of emotional stimuli but also engage users’ sensory, motor, and cognitive systems in ways that resemble real-world embodiment.

At the core of VR-based emotion elicitation lies the principle of immersion. Unlike static or two-dimensional emotional stimuli, immersive VR environments surround the user with multisensory cues that imitate real-world affective situations. When equipped with motion tracking, haptic feedback, and 360-degree visual rendering, VR scenarios can induce fear, joy, sadness, or awe with remarkable intensity and authenticity. A person walking across a narrow bridge suspended over a digital canyon will typically show the same autonomic responses (elevated heart rate, skin conductance spikes, postural adjustments) as they would in a similar physical environment. This immersive realism facilitates not just emotional induction but full-body affective engagement, making VR a uniquely embodied tool for emotion science (Carl et al., 2019).

Researchers have harnessed these affordances to create controlled yet emotionally immersive simulations. For example, VR platforms have been utilised to foster empathy by immersing participants in the embodied perspectives of marginalised groups, such as refugees or individuals with disabilities. These perspective-taking experiences evoke affective resonance and promote pro-social behaviour more effectively than written narratives or video presentations, because users move, look, and act as the digital self (Herrera et al., 2018). The emotional engagement in such cases is not purely cognitive or representational; it arises from embodied simulation within a virtual environment.

AR technologies, although less immersive than VR, offer unique opportunities for contextualised emotional elicitation. AR overlays affective content, such as animated avatars, visual distortions or symbolic elements, onto real-world settings, enhancing emotional significance without disconnecting users from their surroundings. These have been employed in therapeutic contexts, such as exposure therapy for phobias, with augmented reality exposure therapy demonstrating clinical efficacy comparable to in-vivo exposure and a strong therapeutic alliance (Hasan, Alhaj, & Hassoulas, 2023; Eshuis et al., 2021). Because users retain access to their physical bodies and space, AR-based emotion elicitation maintains bodily orientation while seamlessly layering emotional stimuli in meaningful ways.

Another emerging domain is the use of **digital avatars and virtual agents**. These are animated characters capable of facial expressions, gestures, and emotional dialogue, capable of interacting with users in emotionally responsive manners. These agents are increasingly equipped with affect recognition systems that detect and respond to the user’s emotions in real time, creating dynamic affective loops. For example, a virtual therapist may adjust their voice tone and facial expression based on the user’s sadness cues, or a virtual classroom assistant may display encouragement when detecting signs of frustration. Such emotionally contingent responses foster the perception of empathy and relational presence, leading to deeper emotional engagement (Stock-Homburg, 2022; Abdollahi et al., 2022).

Critically, these virtual agents do not simply simulate emotion. They often **elicit** it. Users tend to form emotional bonds with virtual characters, experiencing trust, embarrassment, pride, or compassion depending on the scenario. This socio-emotional interactivity is a cornerstone of embodied digital emotion, as it reflects not just internal states but social dynamics mediated by digital bodies. The design of these agents (including their facial features, voice modulation, eye contact, and physical proximity) significantly shapes the perceived appropriateness and impact (Fischer et al., 2019; Numata et al., 2020).

Despite these promising developments, several challenges complicate the use of digital platforms for embodied emotion elicitation. **One concern is standardisation.** While immersive technologies enable personalisation and contextual richness, this very flexibility makes cross-study comparison and reproducibility more difficult. Unlike film clips or image sets, which have normative emotional ratings, VR environments are often bespoke, and their emotional impact can vary depending on the user’s familiarity with digital media, susceptibility to immersion, or motion sickness.

**Another limitation involves cultural and individual variability.** Emotions elicited through virtual agents may differ across cultures due to varying norms around expression, social hierarchy, and technology interaction. Users with autism spectrum conditions or alexithymia may not respond to emotionally expressive avatars in the same way as neurotypical individuals, highlighting the need for adaptable and inclusive design sensitive to user traits and culture (Park & Lim, 2022; Su et al., 2023).

**Ethical considerations are also paramount.** The capacity to provoke intense emotional responses in virtual environments, especially fear, trauma, or shame, raises concerns about user safety, especially when such responses are triggered unintentionally or without appropriate debriefing. Developers must ensure that emotion-inducing content includes clear warnings, exit mechanisms, and post-exposure emotional regulation tools. As these technologies become more integrated into everyday applications, the need for emotional safeguarding becomes urgent.

**Latency and realism** also remain technical hurdles. Subtle delays in avatar response, unnatural motion patterns, or mismatched vocal tone can produce uncanny valley effects, disrupting emotional immersion. To mitigate this, ongoing research focuses on improving real-time synchronisation between user affect and agent behaviour through predictive modelling and low-latency sensors.

Despite these challenges, the potential for embodied emotional elicitation in digital environments is vast. As sensor technologies, computational models, and immersive design principles evolve, the line between digital and physical emotion continues to blur. In the near future, it is plausible to imagine emotionally intelligent virtual companions that not only recognise our feelings but help us navigate them, while providing comfort, challenge, or motivation based on our embodied emotional profile in real time.

What distinguishes digital emotion elicitation from traditional methods is not just the medium but the nature of the interaction. It is no longer about observing static stimuli but about **inhabiting** a scene, **co-regulating** with digital others, and **embodying** emotion in a multisensory continuum. This shift represents more than a technological enhancement; it redefines how we study, simulate, and perhaps even feel emotion in the digital age.

6.9 Case Studies and Experiments

The theoretical and methodological developments discussed throughout this chapter culminate in a range of case studies and experimental investigations that illustrate how embodied emotion can be effectively elicited, measured, and interpreted using integrated visual and computational strategies. These examples bridge the gap between controlled laboratory protocols and applied emotional experiences in digital, clinical, or social contexts. By examining how priming and provoking stimuli, physiological and behavioural measurement tools, and immersive digital environments coalesce in real-world studies, we gain insight into the practical and scientific value of this multidimensional approach.

Participants were immersed in a virtual public-speaking scenario, a setup known to elicit anxiety and measurable physiological arousal (Pertaub, Slater, & Barker, 2002) in VR audiences. Recent work also validates VR public-speaking training and reports good user acceptance, while multimodal VR studies combine psychophysiology with facial analysis to model affect in real time (Slater et al., 2006; Bachmann, Subramaniam, Born, & Weibel, 2023; Bastida et al., 2024)

An illustrative experiment manipulated posture to test causal effects on emotional responding. Leaning forward (approach-related) versus reclining altered both startle eyeblink magnitude and late positive potential while participants viewed appetitive pictures, indicating that posture can modulate reflexive and electrocortical responses to emotional stimuli (Price, Dieckman, & Harmon-Jones, 2012). Converging work shows that leaning forward increases relative left-frontal cortical activation to appetitive cues, consistent with approach motivation (Harmon-Jones, Gable, & Price, 2011). Recent posture experiments also report small but reliable effects of expansive/upward postures on positive affect and cardiac vagal reactivity (RSA) under controlled lab conditions (Van Cappellen et al., 2022).

In applied digital psychotherapy, conversational agents and affect-aware virtual systems have shown clinically meaningful reductions in depression or anxiety in randomised trials, and reviews now map their opportunities and limits (Fitzpatrick, Darcy, & Vierhile, 2017; Liu et al., 2022; He et al., 2022; Boucher et al., 2021). Systems increasingly fuse vocal/ facial cues with wearable signals to adapt interaction in real time.

In education, affect-aware systems detect facial expressions and eye gaze to adapt content and questions. Classic work such as **Gaze Tutor** dynamically redirects attention using eye-tracking, and **Affective AutoTutor** adjusts strategies based on learners’ affect, yielding engagement benefits and learning gains in certain cohorts. Recent surveys and studies continue to integrate gaze with real-time emotion recognition in learning environments (D’Mello, Olney, Williams, & Hays, 2012; D’Mello & Graesser, 2013; Myers, 2021; Cisse, 2024).

Perhaps the most complex integration of visual, computational, and physiological tools is found in multimodal datasets designed specifically for embodied emotion research. The MAHNOB-HCI dataset, for example, was created through a series of experiments where participants watched emotional film clips while their facial expressions, gaze direction, electrocardiographic activity, respiration, and EEG were recorded simultaneously (Soleymani et al., 2011). This dataset allowed researchers to develop predictive models of emotion that combined both bottom-up and top-down inputs, including context-aware video modelling and affective memory cues. Models trained on such datasets have since been used in healthcare, assistive robotics, and adaptive gaming environments, further emphasising the practical value of multimodal embodied emotion elicitation research.

The WESAD dataset is a widely used benchmark for studying how wearable signals change under different affective states. Participants wore a **RespiBAN** device on the chest (high-rate recording) together with an **Empatica E4** wristband. WESAD provides synchronised streams such as respiration, ECG, EMG, and skin conductance (EDA) from the chest device, and EDA, blood-volume pulse (PPG/BVP), temperature, and 3-axis acceleration from the wristband. Sampling is high on the chest (≈700 Hz) and lower but sufficient for physiology on the wrist (e.g., BVP 64 Hz, EDA/TEMP 4 Hz, ACC 32 Hz). In the original protocol, each subject passes through blocks such as *baseline, stress (Trier Social Stress Test), amusement*, and *meditation*, which are distributed with canonical label codes. For this chapter, we focus on the three most common classes, *Baseline, Stress*, and *Amusement*,because they appear in every subject and are the most frequently modelled across papers.

Figure 8 visualises what these wearable signals look like in practice. Panel A shows a short segment from the chest device: respiration (slow sinusoidal wave), ECG (sharp QRS spikes), EMG (brief muscle bursts), and EDA (slow tonic drift with small phasic responses). Panel B shows the wrist: BVP/PPG (regular pulse peaks), EDA (tonic level), acceleration magnitude (movement), and temperature (slow drift). Together, the two devices provide a compact, multimodal view of physiology that responds to emotional context.

|  |
| --- |
|  |
|  |

Figure 8. Wearable signals in WESAD: (A) chest channels; (B) wrist channels.

WESAD’s structure makes it convenient for **windowed analysis**: we divide each recording into short overlapping windows (e.g., 10 s length, 5 s step), extract robust statistical and spectral features per window, and then learn to distinguish *Baseline* vs. *Stress* vs. *Amusement* using models such as KNN, SVM-RBF, or Random Forest. The subject identifier is always used as a **group** during cross-validation to avoid leakage across people (details in the next subsections). This case study will use WESAD solely as an illustration of the end-to-end process: light preprocessing, quick EDA, subject-aware evaluation, and simple, interpretable diagnostics.

Wearable signals are messy. Wrist acceleration can spike during a gesture. EMG bursts can look like noise. BVP/PPG can momentarily drop. EDA drifts slowly over time. Instead of throwing data away, we apply *gentle*, per-feature cleanup that keeps the median behaviour intact while preventing extreme values from dominating the models.

We keep the preprocessing intentionally simple and leakage-proof. We start by working only with the engineered **numeric** features and leave identifiers such as subject and label untouched. Any infinities that appear from filtering are treated as missing values, so they do not spread through the pipeline. Missing data are then filled with the **median** of the respective feature, with that median calculated only on the **training subjects**. Because wearable streams sometimes produce sudden spikes (a wrist flick in ACC, a brief PPG dropout, a sudden EMG burst), we **winsorise** each feature by trimming its tails at the 1st and 99th percentiles. These cut points are again determined using the training set. Features that are essentially constant across training samples are removed, as near-zero variance provides no discriminative information and only introduces noise. We standardise each feature with **RobustScaler** (centring by the median and scaling by the interquartile range), so the spread is stabilised without being affected by outliers as a z-score would be. All these quantities (medians, percentile cut points, variance mask, scaling factors) are calculated on the training set and then applied unchanged to the held-out test subjects, ensuring a subject-aware pipeline with no information leakage.

To make the effect concrete, Figure 9 contrasts a representative feature before and after preprocessing: the raw histogram shows a long right tail due to motion bursts, whereas the winsorized-and-robust-scaled version preserves the central bulk of the distribution while taming the extremes.

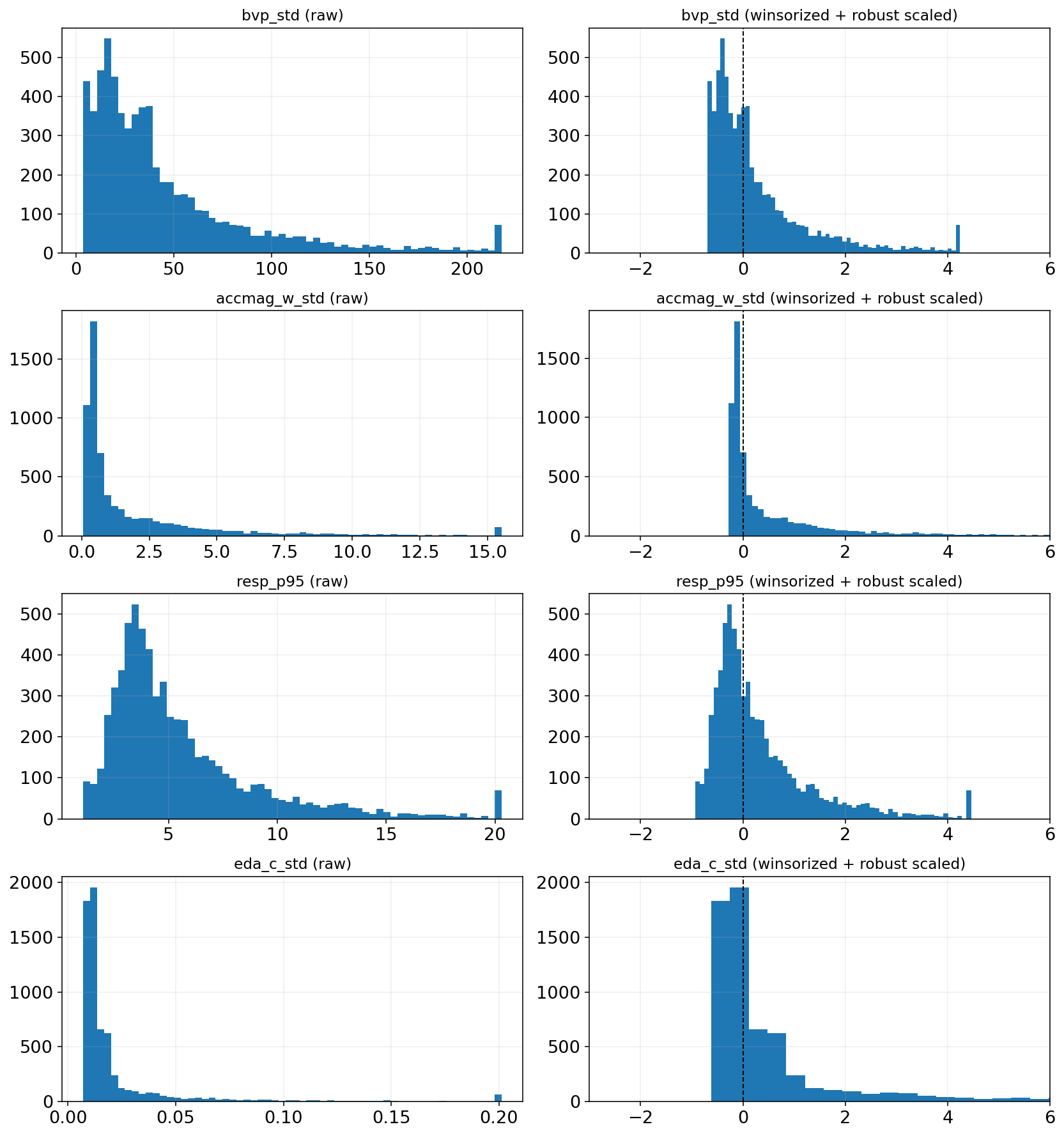


Figure 9. Gentle outlier handling on a single feature: (A) raw histogram; (B) after 1st–99th-percentile clipping and robust scaling (median/IQR)

Before we put any models to work, it helps to look at the shape of the data, and to detect what ranges the features live in, which ones are skewed or heavy-tailed, and where the three WESAD conditions might already separate. Figures 10-12 are designed for a fast but informative pass through the dataset. They are produced after the light preprocessing from the previous step, so you’re looking at the same value space the models will see.

We will start with the marginal histograms in Figure 10. Each small axis shows the distribution of one engineered feature across all subjects and windows. Several sensors naturally produce skewed, long-tailed histograms (e.g., BVP and accelerometer variability), while others are tighter and closer to unimodal (e.g., respiratory percentiles in calm segments). These shapes justify the robust treatment we used: winsorization avoids rare spikes dominating the scale, and robust scaling keeps the bulk of values comparable without forcing strict normality. You’ll also notice that many features have most of their mass near zero after scaling. That’s expected because medians are roughly centred at zero, and the interquartile range defines the unit. This panel is useful for spotting features that still behave oddly (e.g., multimodal or near-constant), which are candidates for later pruning or careful interpretation.

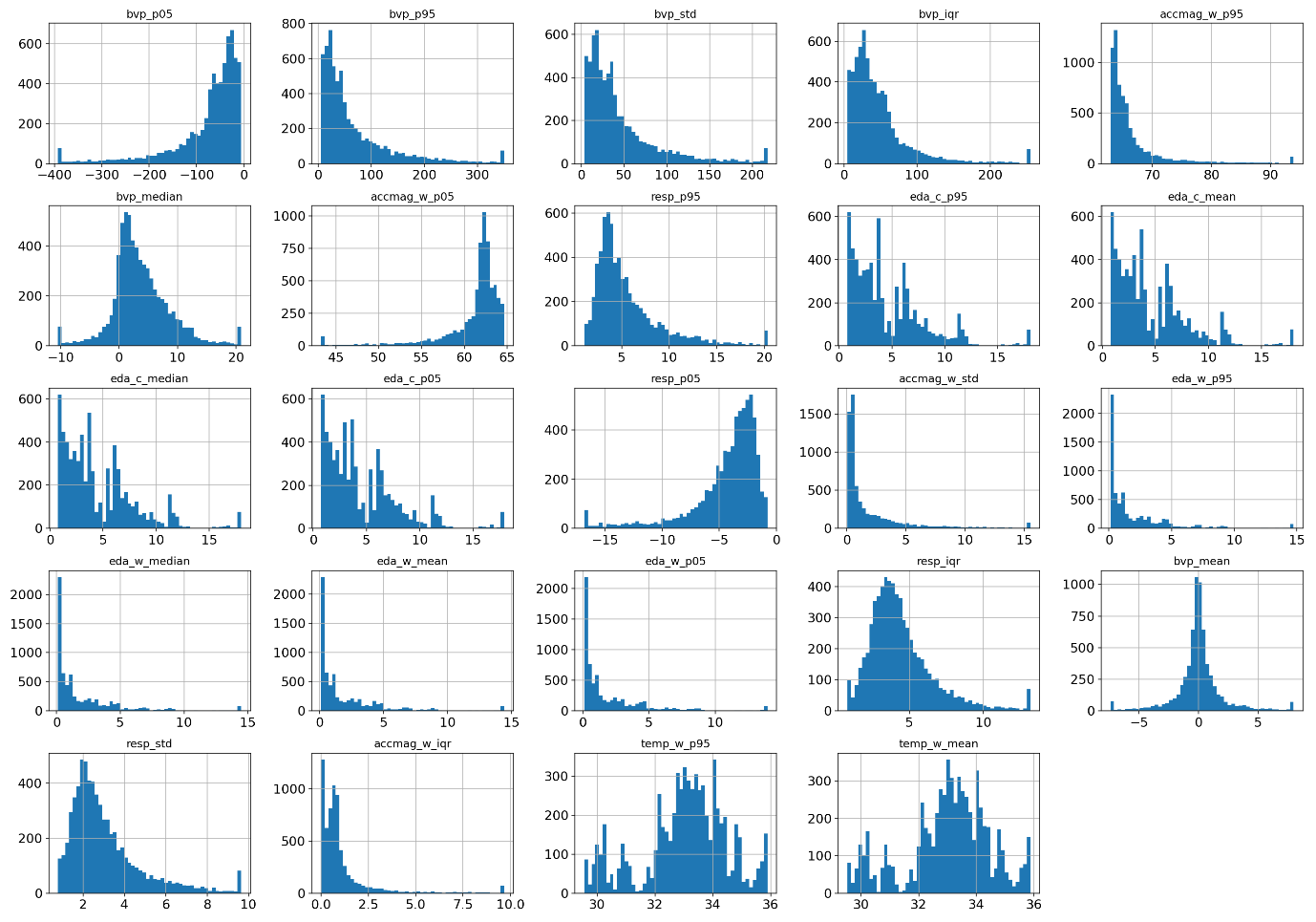


Figure 10. Distribution snapshots of representative engineered features

Next, examine the class-conditioned view in Figure 11. Here, we display the top-K features selected by a straightforward ANOVA F-score (calculated solely on the training split), standardised to z-units, and separated by class. The width of the violin indicates density, while the dashed lines signify the median and quartiles. Several well-known psychophysiological patterns emerge: chest and wrist EDA dispersion and upper percentiles tend to rise during *Stress* compared to *Baseline*, aligning with increased sympathetic activation. Respiration percentiles often increase during stress or active states, whereas *Baseline* shows lower and more compact values. Accelerometer summary statistics often become wider during *Amusement* periods, as participants sometimes move or laugh, whereas *Baseline* generally remains more contained. Not every feature distinguishes clearly. Overlapping violins serve as a helpful reminder that we are dealing with noisy, naturalistic signals.

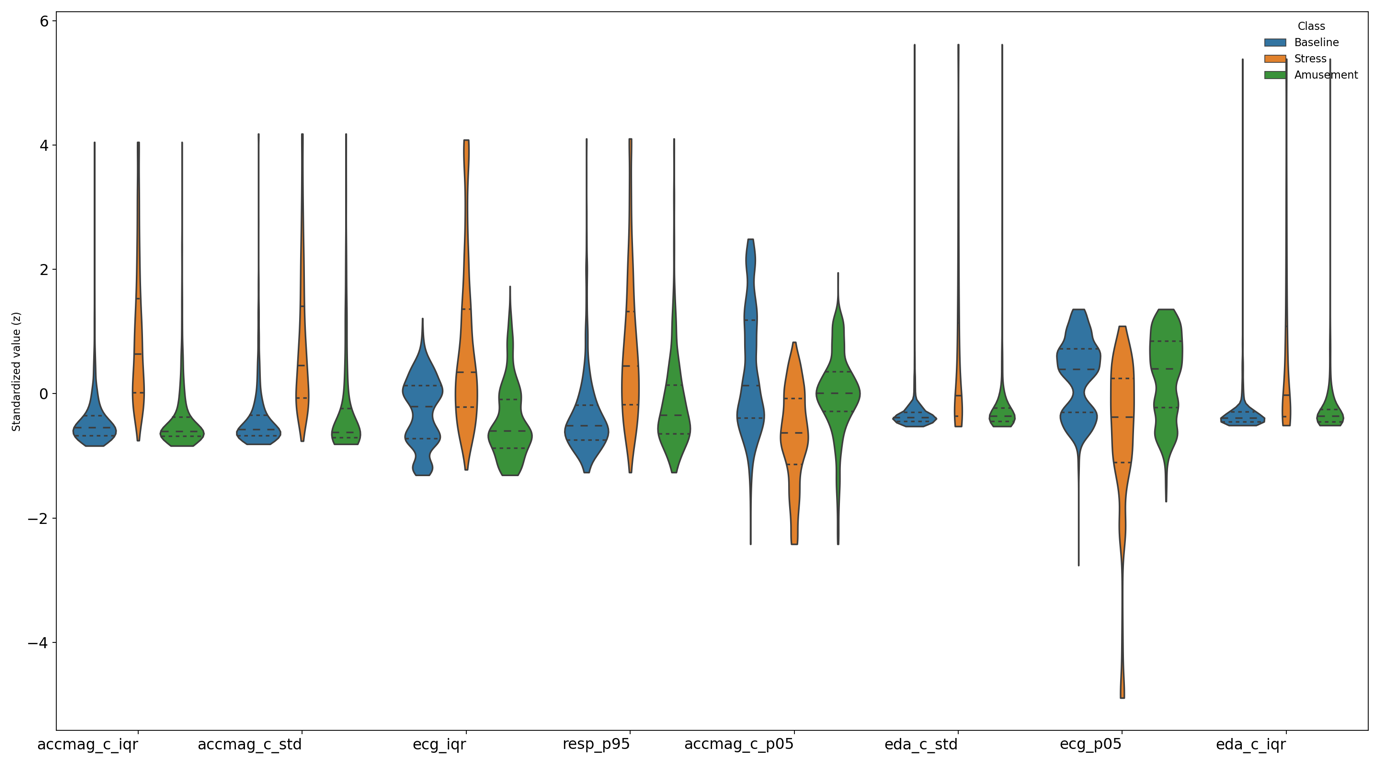


Figure 11. Class-conditioned violin plots for the top-K discriminative features (z-scaled).

Figure 12 offers the same top-K set with box plots, which emphasise nested quantiles deep into the tails. This view is handy when you want to confirm that apparent class differences are not driven only by a handful of extreme windows. For instance, if Stress shows a consistently higher ladder of quantiles for an EDA statistic, that suggests a shift of the bulk, not just a few outliers. Conversely, if only the highest rungs separate, you may be looking at sporadic bursts (e.g., brief motion or sudden SCR flares) rather than a stable level shift. This is a useful diagnostic context when deciding how aggressively to regularise the model or whether to engineer additional robust features.

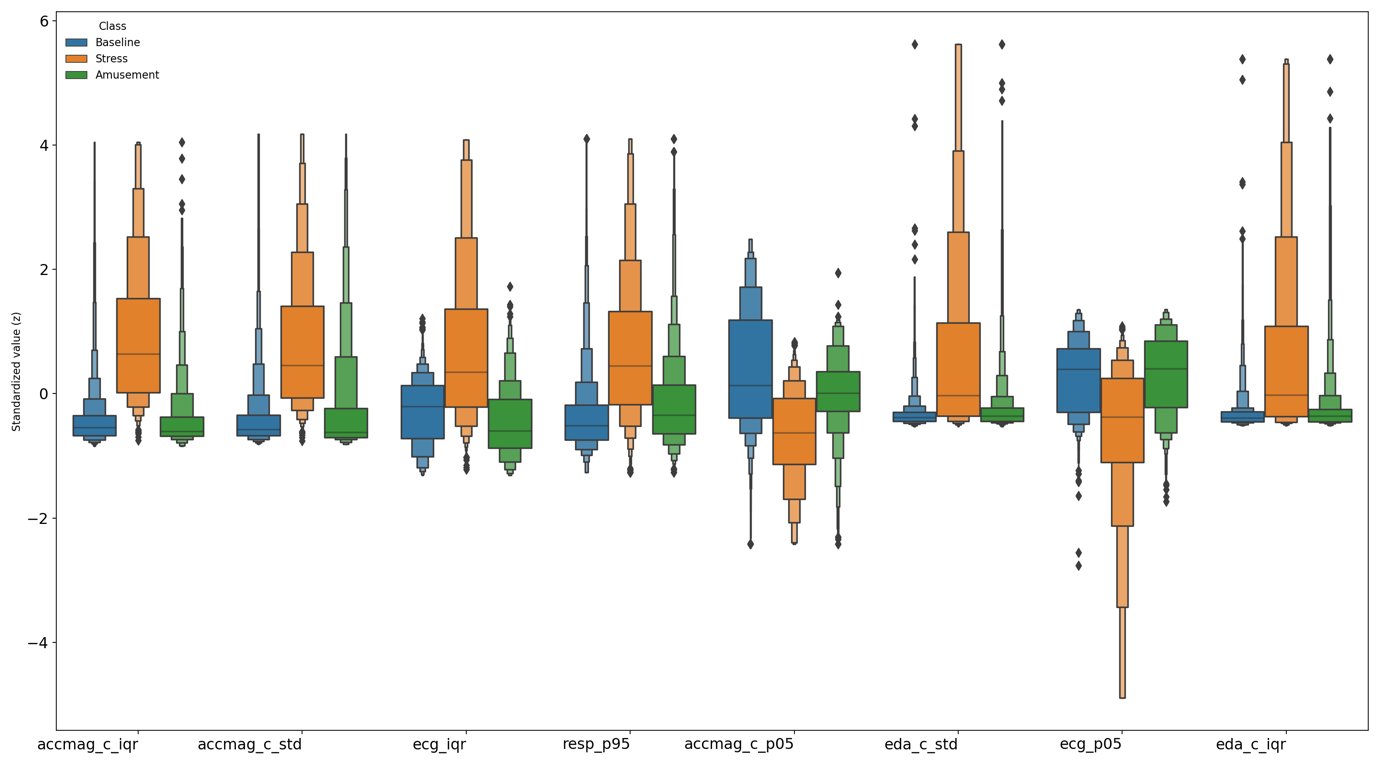
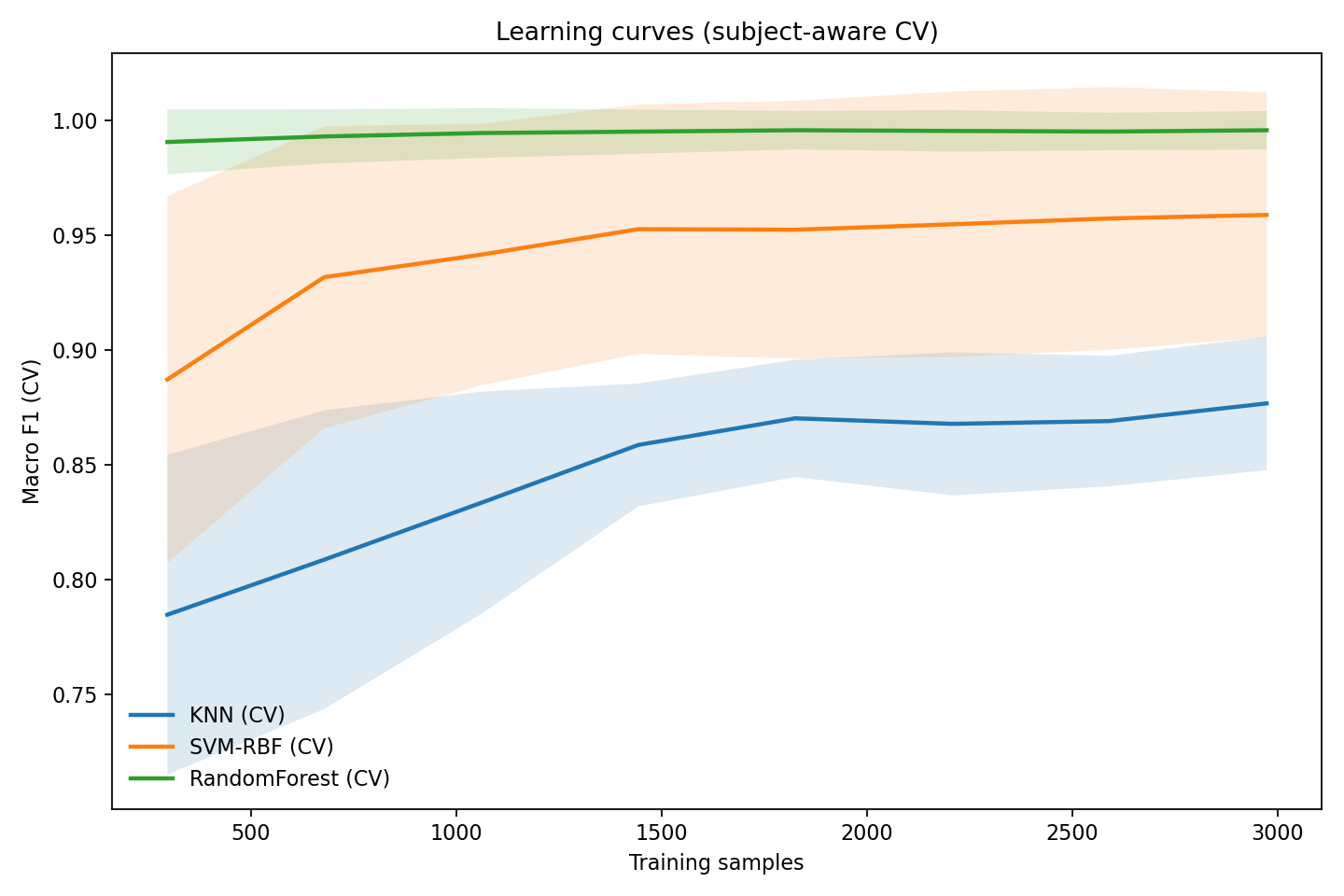


Figure 12. Box plots of the same top-K features, highlighting deep quantiles

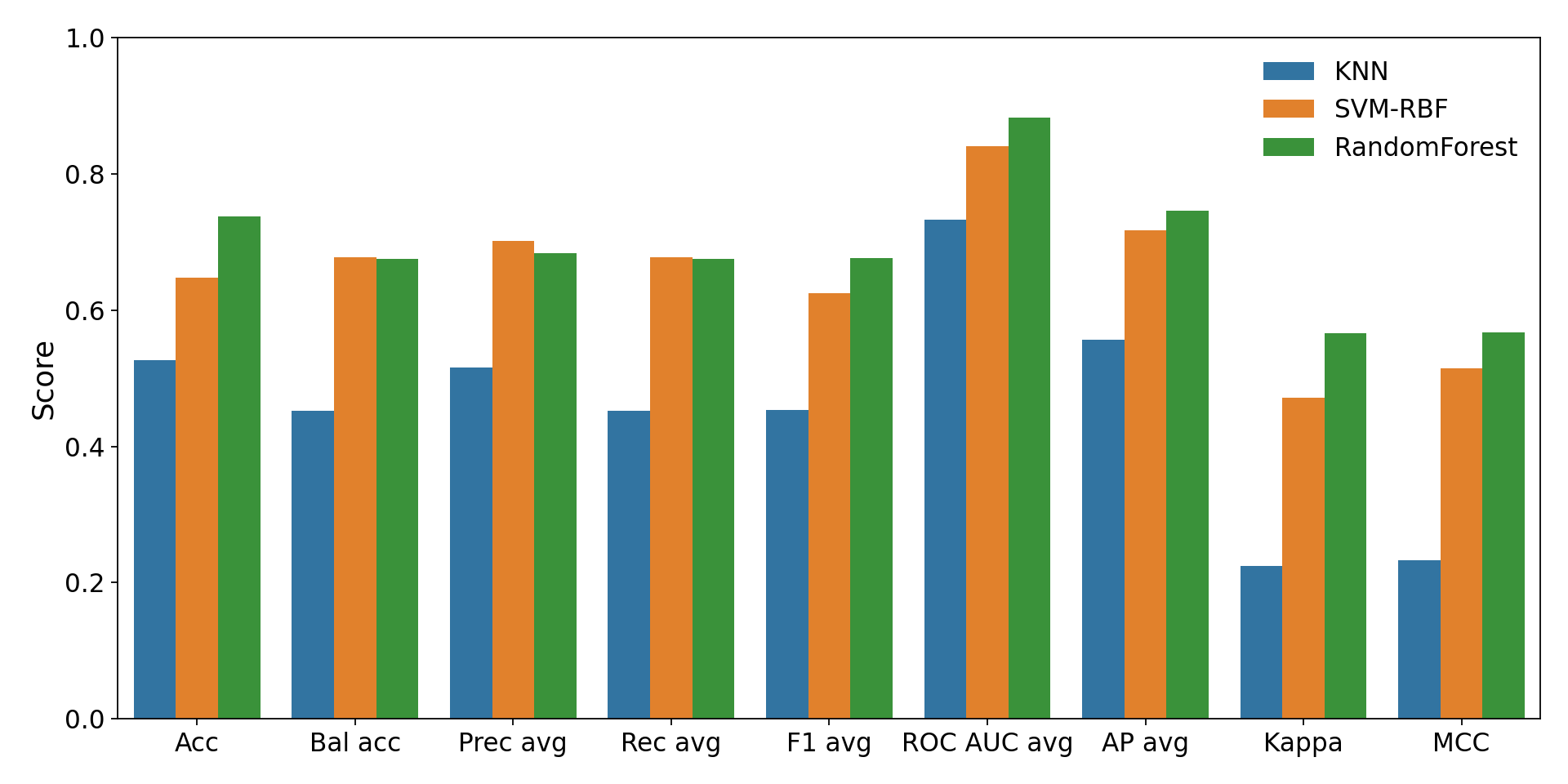
A few notes help when reading these panels. Each distribution combines windows from many participants. Baselines differ by person. To respect that, the modelling section always splits train and test by subject to avoid leakage. All axes use a common standardised scale so features can be compared side by side. The values show distance from the median measured in IQR units, not the original sensor units. When a feature looks bimodal or multimodal, the shape often reflects protocol phases (baseline versus task) and small differences in sensor placement rather than fundamentally different physiology. That is the kind of structure the classifiers will be asked to separate.

We evaluate three complementary classifiers, KNN, SVM with an RBF kernel, and a Random Forest, under a subject-aware split so that people the model sees at training time are never seen at test time. This is the right framing for wearables: generalisation must run across unseen users, not just unseen windows. In Figure 13, we summarise data efficiency with 5-fold cross-validation (CV), where each fold holds out entire subjects so that no subject appears in both training and validation. Each curve shows the mean macro-F1 across the five folds as we increase the number of training windows, and the shaded band is one standard deviation across folds. KNN climbs steadily yet plateaus well below the margin-based methods, which is a sign of underfitting with noisy, heterogeneous physiological features. SVM-RBF reaches a high cross-validated Macro-F1 quickly and then flattens. RandomForest tracks a similar trend but with a consistently higher ceiling, suggesting that ensembles of trees capture nonlinear interactions among respiration, EDA, ECG and accelerometry that a local method like KNN misses. The gap between cross-validated curves and the final test number is expected because the test split represents a distinct set of people. The shape of the curve (rather than the absolute height) is what informs whether adding data is likely to help.



**Figure 13. Learning curves**

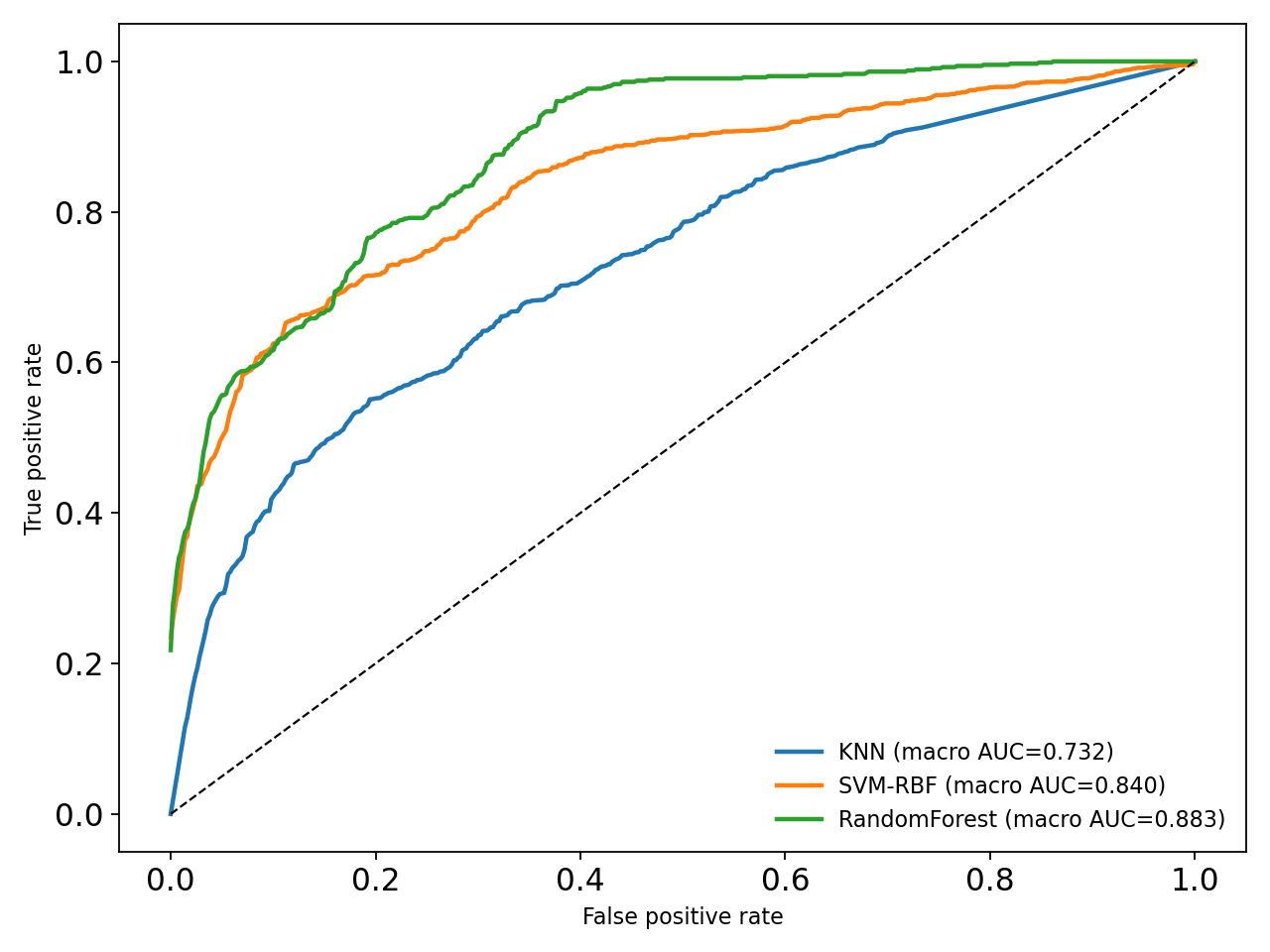
Aggregate scores across several metrics are compactly compared in **Figure 14**. We report a compact set of test-set metrics. Accuracy is the fraction of correct predictions. Balanced accuracy is the average of per-class recalls, which treats the three classes equally. Macro precision, macro recall, and macro F1 average the corresponding per-class scores with equal weight, so no single class dominates the summary. ROC-AUC is computed in a one-vs-rest fashion for each class and then averaged, and Average Precision is the area under the precision–recall curve averaged across classes. We also include two agreement measures that account for chance: Cohen’s kappa and the Matthews correlation coefficient.



**Figure 14. Aggregate metrics by model**

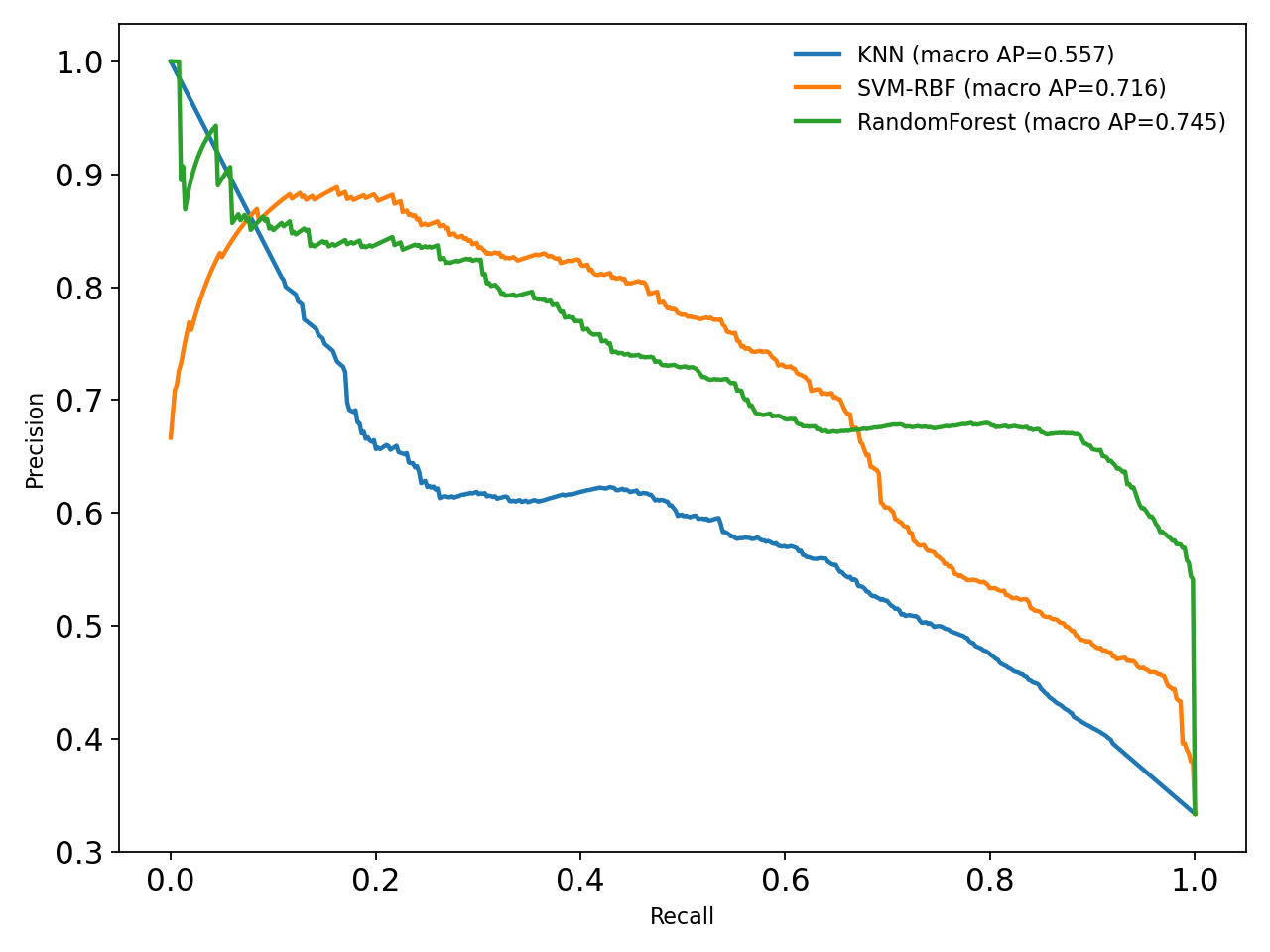
RandomForest leads on Accuracy, Macro F1, ROC-AUC (OvR) and Average Precision (macro), and posts the best agreement measures (Cohen’s κ and MCC). SVM-RBF is a solid second: it edges close to RF on precision-oriented measures but lags on recall for the Stress class. KNN is consistently last across metrics. With window-level features and subject variability, a simple neighbourhood rule struggles to separate Baseline from the activated states. In practice, this means the ensemble not only makes more correct calls overall but also balances performance across *Baseline*, *Stress*, and *Amusement*, whereas KNN struggles to separate the activated states from *Baseline*.

Threshold-free comparisons reinforce this ranking. **Figure 15** shows macro-averaged ROC curves: RandomForest achieves the highest area (AUC ≈ 0.883), SVM-RBF follows (≈ 0.840) and KNN trails (≈ 0.732).



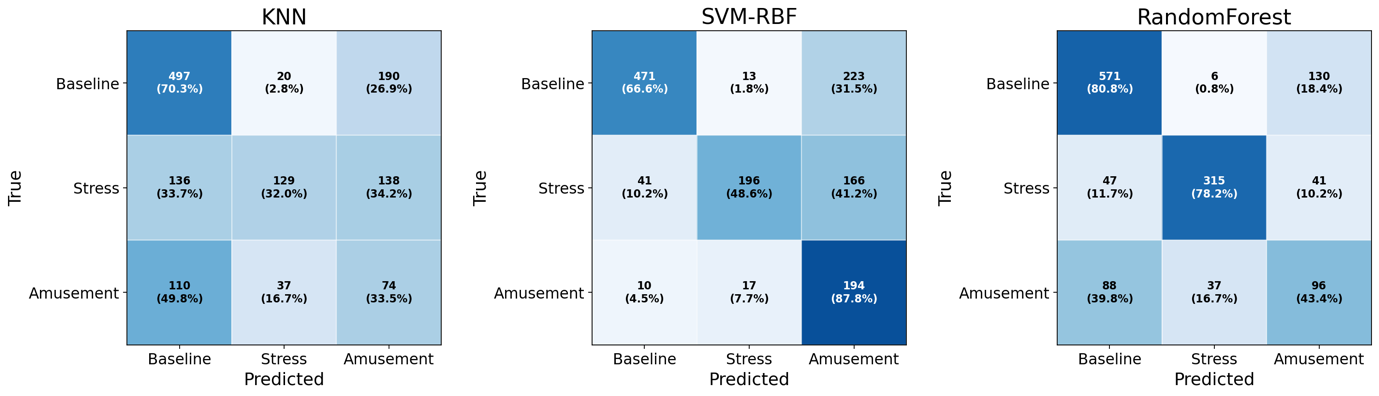
**Figure 15. Macro ROC (one-vs-rest)**

Precision-recall, which is more sensitive to class prevalence, tells the same story in **Figure 16**. RF yields the best macro Average Precision (≈ 0.745), SVM-RBF is next (≈ 0.716), and KNN is far lower (≈ 0.557). Practically, this means RF not only ranks windows correctly more often but also keeps a higher precision at useful recall levels.



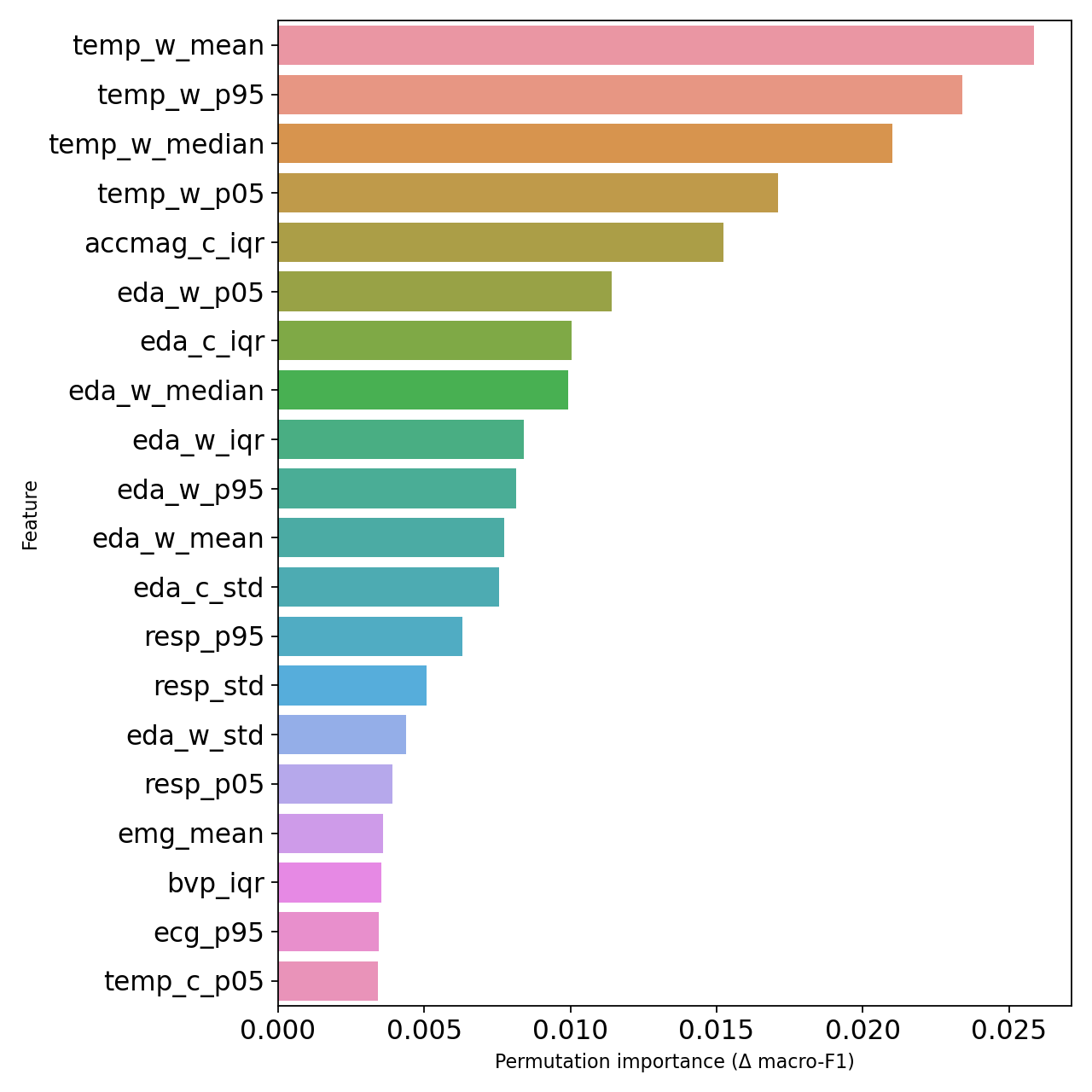
**Figure 16. Macro Precision–Recall**

Where the models succeed and fail is clearest in the confusion matrices of **Figure 17**. RandomForest classifies Stress most reliably (≈ 78% recall) and Baseline strongly (≈ 81%), but Amusement remains the hardest condition, with ≈ 43% recall and a tendency to be absorbed by Baseline. This is not unusual with natural laughter and motion overlapping low-arousal periods. SVM-RBF flips that pattern: it recognises Amusement very well (≈ 88% recall) but over-assigns other states to Amusement, depressing recall for Stress and Baseline. KNN exhibits broad confusion, especially for Stress, which is split almost evenly across the three classes. This matches its lower macro F1.



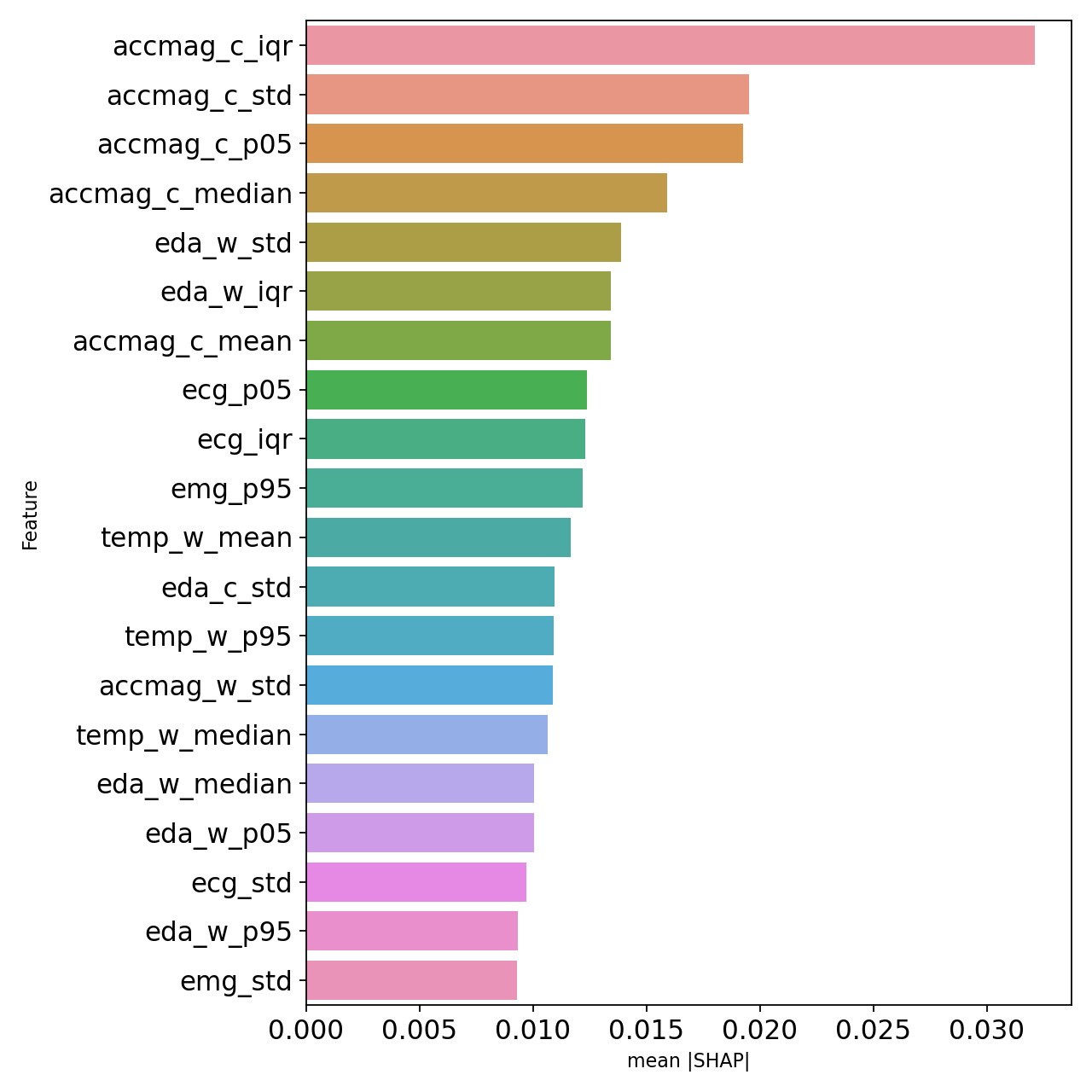
**Figure 17. Confusion matrices**

Finally, we look at why the models decide as they do. **Figure 18** reports permutation importance on the **test** set for the best model (RandomForest): wrist temperature statistics dominate, followed by chest accelerometer variability (IQR) and a family of EDA features from both chest and wrist.



**Figure 18. Permutation importance (RandomForest)**

**Figure 19** presents mean |SHAP| values computed with a tree explainer on the trained forest. Here, chest accelerometer features surface even more prominently, with EDA and ECG quantiles close behind. The two views provide complementary insights rather than identical ones. Permutation importance measures how much each feature affects predictive performance on held-out data. In contrast, SHAP values explain how the model internally attributes importance to its inputs. When features are correlated, such as multiple summaries from temperature or accelerometer channels, their contributions may shift between the two approaches. The physiological story is consistent: motion intensity/variability, skin conductance level dynamics, and peripheral temperature shifts are key drivers separating baseline from aroused states, while ECG quantiles contribute to episodes with marked cardiovascular change.



**Figure 19. SHAP importance (RandomForest)**

Analysis revealed that on subject-held-out WESAD windows, RandomForest is the most reliable all-rounder across metrics and thresholds. SVM-RBF performs competitively and is often the better choice when you care most about recalling *Amusement*. KNN lags on this feature set and is not recommended. The remaining confusion between *Amusement* and *Baseline* points to two practical extensions that are likely to help: add temporal context (either sequence models or features computed over longer horizons) and, where the use case allows it, apply light per-subject calibration. Both are aimed at lifting the *Amusement* recall without giving up *Stress* detection.

Looking beyond the numbers, the case study illustrates a broader shift in method: from isolated snapshots of emotion to a more embodied and context-aware view. The combination of elicitation design, wearable sensing, and subject-aware modelling lets us capture not only what people report but also how they feel, express, regulate, and respond as situations unfold. Emotion is treated less as a static label and more as a dynamic process that emerges from interactions among body, context, and technology.

As this direction develops, a few principles matter. Methods should be transparent and reproducible, especially when multiple data streams are fused. Ethical safeguards belong in the system from the start, particularly when using emotionally charged stimuli or lifelike avatars. And wherever possible, personalisation should guide both stimulus delivery and model interpretation to respect the individual nature of affective experience. With these guardrails in place, integrated approaches like the one shown here can shape how we study and use emotion across scientific, clinical, and social settings.

6.10 Ethical Perspectives and Future Directions

As embodied emotion research advances through increasingly sophisticated methods, ranging from immersive virtual environments and wearable biosensors to AI-driven affect recognition, the importance of grounding this progress in robust ethical frameworks becomes paramount. Emotion, by its very nature, lies at the core of personal identity and vulnerability. The ability to provoke, monitor, and interpret emotional states, especially through embodied and computational systems, raises fundamental questions about consent, autonomy, privacy, and the appropriate boundaries of scientific inquiry and technological design. This concluding section reflects on the ethical dimensions of embodied emotion research and proposes future directions that responsibly harness its potential across scientific and applied domains.

One of the primary ethical concerns arises from the **power asymmetry embedded in emotional technologies**. Unlike traditional experimental setups where participants voluntarily engage in controlled stimuli, many modern platforms, especially those embedded in daily-use devices, can track and respond to emotional signals without explicit awareness. Facial recognition algorithms, passive biometric sensors, and emotion-adaptive agents may infer affective states from micro-expressions or physiological changes that users themselves are not consciously aware of. This passive detection poses a profound risk of emotional surveillance, especially when deployed in educational, workplace, or consumer settings. In such contexts, users must retain clear and ongoing agency over what is collected, how it is interpreted, and to whom the data is accessible (Crawford & Paglen, 2021).

As computational systems are increasingly capable of **modulating emotion**, the line between emotional support and manipulation becomes blurred. While virtual therapists or learning assistants may adjust tone and pacing to reduce distress or improve engagement, these same mechanisms could be employed to nudge user behaviour in commercial or political contexts. The ethics of **emotional persuasion**, particularly through avatars that appear empathetic or human-like, must be critically evaluated. Transparency regarding system intentions, affective capabilities, and data flow should be central design principles in any emotionally responsive system.

**Bias and representation** also demand attention. Emotion recognition systems are often trained on datasets that lack cultural, racial, and neurodiversity. Facial expression data, in particular, has been shown to perform poorly on individuals with darker skin tones, non-Western affective styles, or conditions such as autism spectrum disorder. This not only leads to reduced accuracy but reinforces exclusionary emotional norms, implying that certain expressions are *standard* or *correct* while others are marginalised or misclassified. Future work must prioritise dataset inclusivity and model generalizability, ensuring that emotional systems are valid across diverse populations and do not inadvertently pathologise or ignore authentic emotional expressions (Rhue, 2018).

As emotion elicitation becomes more immersive, the potential for **emotional harm or dysregulation** increases. Provoking fear, shame, or sadness in tightly controlled lab settings may be justifiable for short-term insight, but reproducing these experiences in virtual environments with less supervision can overwhelm users or re-trigger past trauma. Ethical protocols must therefore adapt to account for the **intensity, realism, and duration** of digital affective stimuli. Measures such as emotional safety check-ins, post-session debriefing, and accessible opt-out features should be standard practice in both research and commercial applications.

Looking forward, the future of embodied emotion research lies in **responsible integration**. This means not simply fusing physiological, behavioural, and visual data streams, but doing so in a way that respects the user’s emotional boundaries, preferences, and contexts. Emotion-aware systems should evolve toward personalisation, not only to improve accuracy but also to align with individual emotional goals and sensitivities. For example, affective technologies designed for neurodiverse users may need to accommodate reduced facial expressivity or unique sensory profiles without defaulting to deficit-based assumptions.

Interdisciplinary collaboration will be essential. Ethical frameworks must be co-developed by researchers, ethicists, technologists, designers, and users themselves. Participatory design methods can ensure that emotional systems are not only functional but empowering. This user-centred ethos is particularly important in clinical and educational contexts, where emotional tools should support regulation, growth, and resilience rather than impose normative emotional states.

Another critical direction is the **development of explainable emotional AI**. As models become more opaque and data-hungry, users must be able to understand how emotional predictions are made, on what basis, and with what level of certainty. Explainability is not merely a technical issue. It is a matter of emotional dignity. When systems suggest that someone is angry, anxious, or disengaged, those interpretations must be transparent and open to challenge, not treated as absolute truths.

The future also holds promise in **multisensory emotional modelling**, where touch, voice, motion, and even interoceptive signals (like gut feelings) are integrated into rich embodied models of affect. With wearable technology becoming more sophisticated, the ability to decode these subtle embodied cues will deepen our understanding of how emotions unfold and interact with cognition and behaviour. However, this progress must be guided by **data minimisation and user consent**, ensuring that the richness of emotion is not reduced to a commodity or surveillance metric.

As emotions become increasingly entangled with algorithmic systems, we must ask what kind of emotional future we want to build. Do we seek tools that optimise productivity and compliance, or systems that foster empathy, reflection, and authentic connection? The answer to this question will shape not only the trajectory of research but also the broader cultural meanings we assign to emotion in the digital age. Embodied emotion research, with its ability to trace affect across body, brain, and environment, holds tremendous promise. But its success will not be measured by technical precision alone. It will be defined by its ability to humanise emotion science, to respect the complexity of feeling, and to co-create technologies that serve rather than subvert our emotional lives.

**References:**

Abdollahi, H., Mahoor, M. H., Zandie, R., Siewierski, J., & Qualls, S. H. (2022). Artificial emotional intelligence in socially assistive robots for older adults: a pilot study. *IEEE Transactions on Affective Computing*, *14*(3), 2020-2032.

Aly, L., Godinho, L., Bota, P., Bernardes, G., & da Silva, H. P. (2024). Acting emotions: a comprehensive dataset of elicited emotions. *Scientific data*, *11*(1), 147.

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception*, *33*(6), 717-746.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, *338*(6111), 1225-1229.

Azzalini, D., Rebollo, I., & Tallon-Baudry, C. (2019). Visceral signals shape brain dynamics and cognition. *Trends in cognitive sciences*, *23*(6), 488-509.

Bachmann, M., Subramaniam, A., Born, J., & Weibel, D. (2023). Virtual reality public speaking training: effectiveness and user technology acceptance. *Frontiers in virtual reality*, *4*, 1242544.

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of personality and social psychology*, *71*(2), 230.

Barrett, L. F. (2017). *How emotions are made: The secret life of the brain*. Pan Macmillan.

Barsalou, L. W. (2008). Grounded cognition. *Annu. Rev. Psychol.*, *59*(1), 617-645.

Bastida, L., Sillaurren, S., Loizaga, E., Tomé, E., & Moya, A. (2024). Exploring Human Emotions: A Virtual Reality-Based Experimental Approach Integrating Physiological and Facial Analysis. *Multimodal Technologies and Interaction*, *8*(6), 47.

Baumgartner, T., Esslen, M., & Jäncke, L. From emotion perception to emotion experience. *emotions evoked by pictures and classical music*, *2006*, 60.

Blythe, E., Garrido, L., & Longo, M. R. (2023). Emotion is perceived accurately from isolated body parts, especially hands. *Cognition*, *230*, 105260.

Boğa, M., Koyuncu, M., Kaça, G., & Bayazıt, T. O. (2023). Comparison of emotion elicitation methods: 3 methods, 3 emotions, 3 measures. *Current Psychology*, *42*(22), 18670-18685.

Boucher, E. M., Harake, N. R., Ward, H. E., Stoeckl, S. E., Vargas, J., Minkel, J., ... & Zilca, R. (2021). Artificially intelligent chatbots in digital mental health interventions: a review. *Expert review of medical devices*, *18*(sup1), 37-49.

Capriola-Hall, N. N., Ollendick, T. H., & White, S. W. (2021). Attention deployment to the eye region of emotional faces among adolescents with and without social anxiety disorder. *Cognitive therapy and research*, *45*(3), 456-467.

Carl, E., Stein, A. T., Levihn-Coon, A., Pogue, J. R., Rothbaum, B., Emmelkamp, P., ... & Powers, M. B. (2019). Virtual reality exposure therapy for anxiety and related disorders: A meta-analysis of randomized controlled trials. *Journal of anxiety disorders*, *61*, 27-36.

Carlisle, M., Uchino, B. N., Sanbonmatsu, D. M., Smith, T. W., Cribbet, M. R., Birmingham, W., ... & Vaughn, A. A. (2012). Subliminal activation of social ties moderates cardiovascular reactivity during acute stress. *Health Psychology*, *31*(2), 217.

Carney, D. R., Cuddy, A. J., & Yap, A. J. (2010). Power posing: Brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological science*, *21*(10), 1363-1368.

Cereghetti, D., Coppin, G., Porcherot, C., Cayeux, I., Sander, D., & Delplanque, S. (2024). Beyond self-report measures of arousal: A new priming task to capture activation of relaxing and energizing feelings elicited by odors. *Food Quality and Preference*, *119*, 105227.

Cesario, J., Jonas, K. J., & Carney, D. R. (2017). CRSP special issue on power poses: What was the point and what did we learn?. *Comprehensive Results in Social Psychology*, *2*(1), 1-5.

Cisse, A. H. (2024, November). Real-time Adaptive Learning Environments Using Gaze and Emotion Recognition Engagement and Learning Outcomes. In *International Conference on Computers in Education*.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain sciences*, *36*(3), 181-204.

Coan, J. A., & Allen, J. J. (2004). Frontal EEG asymmetry as a moderator and mediator of emotion. *Biological psychology*, *67*(1-2), 7-50.

Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature reviews neuroscience*, *10*(1), 59-70.

Crawford, K., & Paglen, T. (2021). Excavating AI: The politics of images in machine learning training sets. *Ai & Society*, *36*(4), 1105-1116.

Critchley, H. D., & Garfinkel, S. N. (2017). Interoception and emotion. *Current opinion in psychology*, *17*, 7-14.

D'mello, S., & Graesser, A. (2013). AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *2*(4), 1-39.

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies*, *70*(5), 377-398.

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

De Gelder, B. (2016). *Emotions and the Body*. Oxford University Press.

De Gelder, B., de Borst, A. W., & Watson, R. (2015). The perception of emotion in body expressions. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(2), 149-158.

Dennison, J. (2024). Emotions: functions and significance for attitudes, behaviour, and communication. *Migration studies*, *12*(1), 1-20.

Ekman, P. (1973). Cross-cultural studies of facial expression. *Darwin and facial expression: A century of research in review*, *169222*(1), 45-60.

Ekman, P. (1992). An argument for basic emotions. *Cognition & emotion*, *6*(3-4), 169-200.

Ekman, P., & Friesen, W. V. (1978). Facial action coding system. *Environmental Psychology & Nonverbal Behavior*.

Elgendi, M., Kumar, P., Barbic, S., Howard, N., Abbott, D., & Cichocki, A. (2018). Subliminal priming—state of the art and future perspectives. *Behavioral Sciences*, *8*(6), 54.

Elkjær, E., Mikkelsen, M. B., Michalak, J., Mennin, D. S., & O’Toole, M. S. (2022). Expansive and contractive postures and movement: A systematic review and meta-analysis of the effect of motor displays on affective and behavioral responses. *Perspectives on Psychological Science*, *17*(1), 276-304.

Eshuis, L. V., Van Gelderen, M. J., Van Zuiden, M., Nijdam, M. J., Vermetten, E., Olff, M., & Bakker, A. (2021). Efficacy of immersive PTSD treatments: A systematic review of virtual and augmented reality exposure therapy and a meta-analysis of virtual reality exposure therapy. *Journal of psychiatric research*, *143*, 516-527.

Farahani, H., Watson, P., Bezan, T., Kovač, N., Winter, L. C., Blagojević, M., ... & Jiménez, P. (2024). PsAIchology: An Intelligent Direction in Psychological Sciences. In *10th International Scientific Conference Technics, Informatics and Education-TIE 2024*. Faculty of Technical Sciences Čačak, University of Kragujevac.

Farahani, H., Watson, P., Kovač, N., Sheykhangafshe, F. B., Azadfallah, P., Allahyari, A., ... & Chesli, R. R. (2025). Unraveling the Complexity of Love Addiction Using Machine Learning Algorithms: The Influence of Positive and Negative Affect, Interpersonal Needs, and Self-Hate. In *International Handbook of Love: Transcultural and Transdisciplinary Perspectives* (pp. 1-22). Cham: Springer Nature Switzerland.

Fischer, K., Jung, M., Jensen, L. C., & Aus der Wieschen, M. V. (2019, March). Emotion expression in HRI–when and why. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 29-38). IEEE.

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR mental health*, *4*(2), e7785.

Friesen, W. V. (1972). *Cultural differences in facial expressions in a social situation: An experimental test of the concept of display rules*. University of California, San Francisco.

Galanos, T., Liapis, A., & Yannakakis, G. N. (2021, September). AffectGAN: Affect-based generative art driven by semantics. In *2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW)* (pp. 01-07). IEEE.

Gallese, V., & Sinigaglia, C. (2011). What is so special about embodied simulation?. *Trends in cognitive sciences*, *15*(11), 512-519.

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological inquiry*, *26*(1), 1-26.

Gu, Y., Yang, K., Fu, S., Chen, S., Li, X., & Marsic, I. (2018, July). Multimodal affective analysis using hierarchical attention strategy with word-level alignment. In *Proceedings of the conference. Association for Computational Linguistics. Meeting* (Vol. 2018, p. 2225).

Harmon-Jones, E., Gable, P. A., & Peterson, C. K. (2010). The role of asymmetric frontal cortical activity in emotion-related phenomena: A review and update. *Biological psychology*, *84*(3), 451-462.

Harmon-Jones, E., Gable, P. A., & Price, T. F. (2011). Leaning embodies desire: Evidence that leaning forward increases relative left frontal cortical activation to appetitive stimuli. *Biological psychology*, *87*(2), 311-313.

Hasan, S., Alhaj, H., & Hassoulas, A. (2023). The efficacy and therapeutic alliance of augmented reality exposure therapy in treating adults with phobic disorders: systematic review. *JMIR mental health*, *10*(1), e51318.

He, Y., Yang, L., BE, B. W., BE, S. Z., Qian, C., & BE, T. T. (2022). Mental health chatbot for young adults with depressive symptoms: a single-blind, three-arm, randomized controlled trial. *Journal of Medical Internet Research*, *24*(11).

Herrera, F., Bailenson, J., Weisz, E., Ogle, E., & Zaki, J. (2018). Building long-term empathy: A large-scale comparison of traditional and virtual reality perspective-taking. *PloS one*, *13*(10), e0204494.

Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion. *Journal of Experimental Psychology: General*, *141*(1), 19.

Khalfa, S., Schon, D., Anton, J. L., & Liégeois-Chauvel, C. (2005). Brain regions involved in the recognition of happiness and sadness in music. *Neuroreport*, *16*(18), 1981-1984.

Kim, W., Sung, J., Saakes, D., Huang, C., & Xiong, S. (2021). Ergonomic postural assessment using a new open-source human pose estimation technology (OpenPose). *International Journal of Industrial Ergonomics*, *84*, 103164.

Koelstra, S., Muhl, C., Soleymani, M., Lee, J. S., Yazdani, A., Ebrahimi, T., ... & Patras, I. (2011). Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, *3*(1), 18-31.

Kosti, R., Alvarez, J. M., Recasens, A., & Lapedriza, A. (2019). Context based emotion recognition using emotic dataset. *IEEE transactions on pattern analysis and machine intelligence*, *42*(11), 2755-2766.

Kovač, N., Ratković, K., Farahani, H., & Watson, P. (2024). A practical applications guide to machine learning regression models in psychology with Python. *Methods in Psychology*, *11*, 100156.

Kovač, N., Ratković, K., Farahani, H., & Watson, P. (2025b). Machine learning regression models for internal shame. Acta Psychologica, 260, 105721.

Kovač, N., Ratković, K., Watson, P., Farahani, H., & Bagheri Sheykhangafshe, F. (2025). Machine learning classification models for predicting chronic pain. *Current Psychology*, 1-14.

Krumhuber, E. G., Küster, D., Namba, S., & Skora, L. (2021). Human and machine validation of 14 databases of dynamic facial expressions. *Behavior research methods*, *53*(2), 686-701.

Kurdi, B., Lozano, S., & Banaji, M. R. (2017). Introducing the open affective standardized image set (OASIS). *Behavior research methods*, *49*(2), 457-470.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*(39-58), 3.

LeBlanc, V. R., Brazil, V., & Posner, G. D. (2024). More than a feeling: emotional regulation strategies for simulation-based education. *Advances in Simulation*, *9*(1), 53.

Lee, J., Kim, S., Kim, S., Park, J., & Sohn, K. (2019). Context-aware emotion recognition networks. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10143-10152).

Li, S., Deng, W., & Du, J. (2017). Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2852-2861).

Li, W., Moallem, I., Paller, K. A., & Gottfried, J. A. (2007). Subliminal smells can guide social preferences. *Psychological science*, *18*(12), 1044-1049.

Liu, H., Peng, H., Song, X., Xu, C., & Zhang, M. (2022). Using AI chatbots to provide self-help depression interventions for university students: A randomized trial of effectiveness. *Internet Interventions*, *27*, 100495.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of cognition*, *1*(1), 16.

Matsumoto, D., Keltner, D., Shiota, M. N., O’Sullivan, M. A. U. R. E. E. N., & Frank, M. (2008). Facial expressions of emotion. *Handbook of emotions*, *3*, 211-234.

Mikels, J. A., Fredrickson, B. L., Larkin, G. R., Lindberg, C. M., Maglio, S. J., & Reuter-Lorenz, P. A. (2005). Emotional category data on images from the International Affective Picture System. *Behavior research methods*, *37*(4), 626-630.

Miranda-Correa, J. A., Abadi, M. K., Sebe, N., & Patras, I. (2018). Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE transactions on affective computing*, *12*(2), 479-493.

Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2020, April). M3er: Multiplicative multimodal emotion recognition using facial, textual, and speech cues. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 02, pp. 1359-1367).

Mohammad, S. M. (2022). Ethics sheet for automatic emotion recognition and sentiment analysis. *Computational Linguistics*, *48*(2), 239-278.

Mollahosseini, A., Hasani, B., & Mahoor, M. H. (2017). Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, *10*(1), 18-31.

Mousavi, S. M. H., Besenzoni, M., Andreoletti, D., Peternier, A., & Giordano, S. (2023, September). The Magic XRoom: A Flexible VR Platform for Controlled Emotion Elicitation and Recognition. In *Proceedings of the 25th International Conference on Mobile Human-Computer Interaction* (pp. 1-5).

Myers, M. H. (2021). Automatic detection of a student’s affective states for intelligent teaching systems. *Brain Sciences*, *11*(3), 331.

Nandi, A., Xhafa, F., Subirats, L., & Fort, S. (2021). Real-time emotion classification using eeg data stream in e-learning contexts. *Sensors*, *21*(5), 1589.

Nava, E., & Turati, C. (2020). Subliminal affective priming changes the ‘feeling’towards neutral objects in infancy. *Social neuroscience*, *15*(4), 447-457.

Niedenthal, P. M. (2007). Embodying emotion. *Science*, *316*(5827), 1002-1005.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and social psychology review*, *9*(3), 184-211.

Numata, T., Sato, H., Asa, Y., Koike, T., Miyata, K., Nakagawa, E., ... & Sadato, N. (2020). Achieving affective human–virtual agent communication by enabling virtual agents to imitate positive expressions. *Scientific reports*, *10*(1), 5977.

Park, J., & Han, G. (2018). Collectivism and the development of indigenous psychology in South Korea. In *Asia-Pacific perspectives on intercultural psychology* (pp. 53-74). Routledge.

Park, S., & Whang, M. (2022). Empathy in human–robot interaction: Designing for social robots. *International journal of environmental research and public health*, *19*(3), 1889.

Pertaub, D. P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence*, *11*(1), 68-78.

Pessoa, L. (2013). *The cognitive-emotional brain: From interactions to integration.*

Polo, E. M., Iacomi, F., Rey, A. V., Ferraris, D., Paglialonga, A., & Barbieri, R. (2025). Advancing emotion recognition with Virtual Reality: A multimodal approach using physiological signals and machine learning. *Computers in Biology and Medicine*, *193*, 110310.

Price, T. F., & Harmon‐Jones, E. (2015). Embodied emotion: The influence of manipulated facial and bodily states on emotive responses. *Wiley Interdisciplinary Reviews: Cognitive Science*, *6*(6), 461-473.

Price, T. F., Dieckman, L. W., & Harmon-Jones, E. (2012). Embodying approach motivation: Body posture influences startle eyeblink and event-related potential responses to appetitive stimuli. *Biological Psychology*, *90*(3), 211-217.

Reddan, M. C., Chang, L., Kragel, P., & Wager, T. D. (2024). Somatosensory and motor contributions to emotion representation. *arXiv preprint arXiv:2411.08973*.

Rhue, L. (2018). Racial influence on automated perceptions of emotions. *Available at SSRN 3281765*.

Rohr, M., & Wentura, D. (2021). Degree and complexity of non-conscious emotional information processing–a review of masked priming studies. *Frontiers in human neuroscience*, *15*, 689369.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161.

Salamone, P. C., Legaz, A., Sedeño, L., Moguilner, S., Fraile-Vazquez, M., Campo, C. G., ... & Ibañez, A. (2021). Interoception primes emotional processing: multimodal evidence from neurodegeneration. *Journal of Neuroscience*, *41*(19), 4276-4292.

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and emotion*, *24*(7), 1153-1172.

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health*, *5*, 258.

Shu, L., Xie, J., Yang, M., Li, Z., Li, Z., Liao, D., ... & Yang, X. (2018). A review of emotion recognition using physiological signals. *Sensors*, *18*(7), 2074.

Slater, M., Pertaub, D. P., Barker, C., & Clark, D. M. (2006). An experimental study on fear of public speaking using a virtual environment. *CyberPsychology & Behavior*, *9*(5), 627-633.

Smith, R., Thayer, J. F., Khalsa, S. S., & Lane, R. D. (2017). The hierarchical basis of neurovisceral integration. *Neuroscience & biobehavioral reviews*, *75*, 274-296.

Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2011). A multimodal database for affect recognition and implicit tagging. *IEEE transactions on affective computing*, *3*(1), 42-55.

Somarathna, R., Bednarz, T., & Mohammadi, G. (2022). Virtual reality for emotion elicitation–a review. *IEEE Transactions on Affective Computing*, *14*(4), 2626-2645.

Stepper, S., & Strack, F. (1993). Proprioceptive determinants of emotional and nonemotional feelings. *Journal of personality and social psychology*, *64*(2), 211.

Stock-Homburg, R. (2022). Survey of emotions in human–robot interactions: Perspectives from robotic psychology on 20 years of research. *International Journal of Social Robotics*, *14*(2), 389-411.

Su, H., Qi, W., Chen, J., Yang, C., Sandoval, J., & Laribi, M. A. (2023). Recent advancements in multimodal human–robot interaction. *Frontiers in Neurorobotics*, *17*, 1084000.

Torre, I., Carrigan, E., McDonnell, R., Domijan, K., McCabe, K., & Harte, N. (2019, October). The effect of multimodal emotional expression and agent appearance on trust in human-agent interaction. In *Proceedings of the 12th ACM SIGGRAPH conference on motion, interaction and games* (pp. 1-6).

Tsiourti, C., Weiss, A., Wac, K., & Vincze, M. (2019). Multimodal integration of emotional signals from voice, body, and context: Effects of (in) congruence on emotion recognition and attitudes towards robots. *International Journal of Social Robotics*, *11*(4), 555-573.

Tzirakis, P., Chen, J., Zafeiriou, S., & Schuller, B. (2021). End-to-end multimodal affect recognition in real-world environments. *Information Fusion*, *68*, 46-53.

Umair, M., Chalabianloo, N., Sas, C., & Ersoy, C. (2021). HRV and stress: A mixed-methods approach for comparison of wearable heart rate sensors for biofeedback. *IEEe Access*, *9*, 14005-14024.

Van Cappellen, P., Ladd, K. L., Cassidy, S., Edwards, M. E., & Fredrickson, B. L. (2022). Bodily feedback: expansive and upward posture facilitates the experience of positive affect. *Cognition and Emotion*, *36*(7), 1327-1342.

Van der Wel, P., & Van Steenbergen, H. (2018). Pupil dilation as an index of effort in cognitive control tasks: A review. *Psychonomic bulletin & review*, *25*(6), 2005-2015.

Vetter, P., Badde, S., Phelps, E. A., & Carrasco, M. (2019). Emotional faces guide the eyes in the absence of awareness. *ELife*, *8*, e43467.

Visalli, A., Ambrosini, E., Viviani, G., Sambataro, F., Tenconi, E., & Vallesi, A. (2023). On the relationship between emotions and cognitive control: Evidence from an observational study on emotional priming Stroop task. *Plos one*, *18*(11), e0294957.

Wood, A., Rychlowska, M., Korb, S., & Niedenthal, P. (2016). Fashioning the face: sensorimotor simulation contributes to facial expression recognition. *Trends in cognitive sciences*, *20*(3), 227-240.

Yan, W. J., Wu, Q., Liang, J., Chen, Y. H., & Fu, X. (2013). How fast are the leaked facial expressions: The duration of micro-expressions. *Journal of nonverbal behavior*, *37*(4), 217-230.

Zhang, Z. (2012). Microsoft kinect sensor and its effect. *IEEE multimedia*, *19*(2), 4-10.