uOttawa

Real-Time Player Engagement Measurement In Video Games

By

Ammar Rasid

Thesis submitted to the University of Ottawa
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Computer Science

School of Electrical Engineering and Computer Science
Faculty of Engineering
University of Ottawa

Examining Committee

The following served on the Examining Committee for this thesis.

Internal Member: Miodrag Bolic
Professor,
School of Electrical Engineering and Computer Science
University of Ottawa

Internal Member: Hussein Al Osman
Associate Professor,
School of Electrical Engineering and Computer Science
University of Ottawa

Internal Member: Majid Komeili
Associate Professor,
School of Computer Science
Carleton University

External Examiner: Jianchuan Liu
Professor,
School of Computing Science
Simon Fraser University

Supervisor(s): Shervin Shirmohammadi
Professor,
School of Electrical Engineering and Computer Science
University of Ottawa

Declaration

I hereby certify that this thesis is entirely my own original work except where otherwise indicated. I am aware of the University's regulations concerning plagiarism, including those concerning consequent disciplinary actions. Any use of the works of any other author, in any form, is properly acknowledged at their point of use.

Abstract

Player engagement is crucial for video games, directly impacting satisfaction, retention, and commercial success. Game developers currently rely on post-hoc analytics or sales metrics that cannot capture real-time engagement fluctuations, while research approaches depend on intrusive methods requiring specialized equipment, creating a gap between practical needs and current capabilities.

This thesis investigates non-intrusive player engagement measurement methods for both game developers seeking practical optimization tools and researchers studying engagement dynamics. It identifies Flow Theory—which posits optimal engagement occurs when skill and challenge are balanced—as a promising framework for real-time prediction.

The MultiPENG study evaluated engagement across multiple modalities, revealing human judges achieved only 50% accuracy with poor inter-rater agreement (Krippendorff's $\alpha = 0.04$-0.09). Computational approaches demonstrated effective performance, with facial footage (63% accuracy), EEG signals (61%), and eye metrics (59%) showing that the webcam-based approach offered the best balance between performance and practicality. Most significantly, a model using only player skill and challenge as predictors (67% accuracy) performed on par with complex multimodal approaches (65% accuracy), empirically validating Flow Theory. Despite these promising results, the observed "cold-start" sensitivity suggests careful interpretation when generalizing to new participants.

Building on these insights, a novel telemetry-based framework was developed using PlayerUnknown's Battlegrounds—a challenging case study selected for its complex environment combining shooting, combat, scavenging, survival mechanics, and large-scale multiplayer interactions across diverse gameplay phases. The framework's hybrid architecture combining Graph Convolutional Networks with Transformers outperformed Transformer-only models (73% vs. 67% accuracy). Requiring just one minute of gameplay data, the system can proactively forecast engagement by estimating skill and challenge in future timesteps, and then mapping them to an engagement level. Performance matches questionnaire-based methods while operating non-intrusively with standard game telemetry.

The primary contributions include: (1) the MultiPENG dataset with synchronized data across modalities enabling direct comparison; (2) empirical validation of Flow Theory, demonstrating skill-challenge metrics can match complex multimodal approaches; (3) a methodology for measuring skill and challenge directly from gameplay telemetry; and (4) a real-time engagement framework combining Graph Convolutional Networks with Transformers that enables adaptive experiences without specialized equipment. These contributions serve game developers seeking player experience optimization tools and researchers investigating engagement dynamics in interactive systems.

Acknowledgements

I would like to express my sincere gratitude to the University of Ottawa for their generous financial support through the International Doctoral Scholarship and Admission Scholarship, which made my doctoral studies possible. I also gratefully acknowledge the funding received for this research project from Advanced Micro Devices (AMD) through the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant number ALLRP556311-20. This financial support was instrumental in enabling the successful completion of my research work.

My utmost gratitude to my supervisor, Prof. Shervin Shirmohammadi, who has been the academic equivalent of a father figure, a role model, and a supportive friend before he is my supervisor. My deepest thanks to Prof. Mohamed Hefeeda of Simon Fraser University for supporting and directing me throughout my PhD journey. I would like to extend my thanks to Prof. Ihab Amer for his support during my time working with AMD, Dr. Khaled Diab for his support at the beginning of this project, and Dr. Kareem Darwish of QCRI for his support and guidance since my Masters.

I cannot express how grateful I am to my family, and especially my dear wife and partner for her relentless support throughout this arduous journey.

And above all, all praise be to God.

Contents

List of Figures

List of Tables

List of Abbreviations

AR	Augmented Reality
AU	Action Unit
AUC	Area Under Curve
DAU	Daily Active Users
ECG	Electrocardiogram
EDA	Electrodermal Activity
EEG	Electroencephalography
EFM	Experience Fluctuation Model
EMG	Electromyography
ESM	Experience Sampling Method
FAU	Facial Action Unit
FACS	Facial Action Coding System
GCN	Graph Convolutional Network
GEQ	Game Experience Questionnaire
GSR	Galvanic Skin Response
HR	Heart Rate
HRV	Heart Rate Variability
IEQ	Immersive Experience Questionnaire
KPI	Key Performance Indicator
MAU	Monthly Active Users
MultiPENG	Multimodal Player Engagement
PANAS	Positive and Negative Affect Schedule
PENS	Player Experience of Need Satisfaction
PEQ	Player Engagement Questionnaire
PPG	Photoplethysmography
PSD	Power Spectral Density
PUBG	PlayerUnknown's Battlegrounds
QRE	Questionnaire
RMSSD	Root Mean Square of Successive Differences
ROC	Receiver Operating Characteristic
SDRR	Standard Deviation of RR Intervals
SEM	Standard Error of the Mean
SFV	Street Fighter V
STD	Standard Deviation
UGC	User-Generated Content
VR	Virtual Reality

Chapter 1

Introduction

1.1 Background and Motivation

The video gaming industry has undergone remarkable growth in recent years, now generating more revenue than the music and movie industries combined [1]. This economic success stems from gaming's unique position as an interactive medium, where player engagement directly influences satisfaction, retention, and ultimately, commercial performance. As game developers strive to "out-fun" competitors [1], measuring and enhancing player engagement has become crucial for success in this highly competitive landscape.

Player engagement represents a multidimensional construct encompassing cognitive, emotional, and behavioral aspects of the player experience [2]. It manifests differently across various gaming contexts—from immersive single-player narratives to dynamic multiplayer environments—and fluctuates throughout gameplay sessions. While the industry employs various metrics to track engagement indirectly (Daily Active Users, retention rates, session length), capturing the psychological state of engagement in real-time remains challenging.

Traditional approaches to measuring player engagement fall into two main categories, each with significant limitations. Post-game questionnaires like the Game Experience Questionnaire (GEQ) [3] provide comprehensive assessments but suffer from recall bias and cannot capture moment-to-moment fluctuations [4]. Physiological measurements (heart rate, skin conductance, EEG) offer continuous objective data but require specialized equipment, create artificial gaming conditions, and typically involve complex signal processing that prevents real-time detection [5], [6].

The increasing scale and complexity of modern games further complicates engagement measurement. Contemporary games feature vast open worlds, intricate progression systems, and sophisticated multiplayer interactions—all elements that influence player engagement in nuanced ways. As gaming experiences grow more diverse and personalized, developing robust, non-intrusive methods for real-time engagement measurement becomes increasingly important for both academic research and industry applications.

1.2 Research Problem

This thesis addresses a fundamental challenge in video game research and development: how to measure player engagement in real-time without disrupting the gameplay experience. While numerous studies have explored engagement through various lenses—from psychological constructs to physiological correlates—a significant gap exists between theoretical understanding and practical implementation. Specifically, this research tackles the following key problems:

1. The lack of comprehensive comparison between different modalities for engagement measurement under identical gaming conditions, making it difficult to determine which approaches offer the best balance of accuracy and practicality

2. The absence of reliable non-intrusive methods for real-time engagement estimation in complex gaming environments, particularly methods that can operate without specialized equipment or gameplay interruption

3. The challenge of translating theoretical engagement frameworks (particularly Flow Theory) into practical measurement systems that can function across diverse game genres and player demographics

These problems represent significant barriers to understanding and enhancing player experiences in modern games. Without reliable real-time engagement measurement, developers cannot implement dynamic adaptations or identify precise moments where player interest wanes. Similarly, researchers lack the tools to study engagement fluctuations during natural gameplay, limiting our understanding of this complex psychological construct.

1.3 Theoretical Foundation

This thesis builds upon Flow Theory [7] as its primary theoretical framework. Flow Theory posits that optimal engagement occurs when an activity's challenge level matches the individual's skill level. When applied to gaming, this framework suggests that players experience peak engagement when game difficulty aligns with their abilities—neither so challenging that it causes frustration nor so simple that it leads to boredom.

Flow Theory offers several advantages for engagement measurement. First, it provides a clear operational definition through the skill-challenge balance, transforming an abstract psychological state into measurable components. Second, it aligns with industry practices, where dynamic difficulty adjustment and matchmaking systems already aim to maintain optimal challenge levels. Third, it offers a parsimonious approach to engagement estimation, potentially capturing the essence of player engagement with fewer variables than complex multimodal systems.

While numerous engagement frameworks exist in the literature, this thesis specifically examines whether Flow Theory's relatively simple premise can outperform or complement more complex measurement approaches. By investigating how skill and challenge can be estimated from gameplay data, and how these

estimates relate to self-reported engagement, this research aims to validate Flow Theory as a practical foundation for real-time engagement measurement in gaming contexts.

1.4 Research Questions and Objectives

This thesis addresses the following research questions:

1. Which modalities and measurement approaches most effectively capture player engagement in real-world gaming environments?
2. Can Flow Theory's skill-challenge framework provide an effective foundation for non-intrusive, real-time engagement measurement?
3. How can standard game telemetry data be transformed into meaningful engagement metrics without requiring specialized equipment or gameplay interruption?
4. To what extent can engagement measurement approaches generalize across different game genres and player demographics?

To address these questions, the research pursues three primary objectives:

1. To create a comprehensive multimodal dataset enabling direct comparison between different sensing approaches for engagement measurement
2. To evaluate the relative performance of different measurement approaches, particularly comparing complex multimodal systems to simpler Flow Theory-based metrics
3. To develop and validate a framework for real-time, non-intrusive engagement measurement using standard game telemetry data

These objectives align with the broader research directions identified in the literature: developing context-aware algorithms that can function across diverse gaming contexts, exploring methods to capture temporal dynamics without disrupting gameplay, and creating computationally efficient approaches that process data in real-time for practical applications.

1.5 Methodology and Approach

This thesis employs a progressive research methodology that moves from theoretical exploration to practical implementation:

1. **Comprehensive Literature Review**: Examining existing approaches to player engagement measurement across various modalities, ground truth collection methods, and modeling techniques to identify research gaps and establish theoretical understanding

2. **Multimodal Experimental Investigation**: Conducting a controlled experiment (MultiPENG) collecting synchronized data across multiple modalities while implementing a non-intrusive approach to obtaining ground truth engagement measurements

3. **Telemetry-Based Framework Development**: Building on insights from the experimental investigation to develop a practical framework for real-time engagement measurement using game telemetry data

4. **Cross-Domain Validation**: Testing the developed framework across different game genres to assess its generalizability and practical applicability

This approach bridges the gap between theoretical understanding and practical application, addressing the identified research questions through both controlled experimentation and real-world implementation. By progressing from comprehensive review to multimodal investigation and finally to telemetry-based measurement, the research provides a systematic exploration of player engagement estimation in modern gaming environments.

1.6 Thesis Contributions

This thesis makes several significant contributions to the field of player engagement measurement in video games:

1. **Comprehensive Multimodal Dataset**: The MultiPENG dataset provides synchronized engagement data across multiple modalities (webcam footage, gameplay footage, EEG, eye tracking, heart rate, controller inputs), enabling direct comparison of their effectiveness in identical gaming contexts

2. **Innovative Ground Truth Collection**: An Experience Sampling Method implemented during natural gameplay pauses minimizes both gameplay disruption and recall bias, offering a more ecologically valid approach to engagement measurement

3. **Empirical Validation of Flow Theory**: Experimental evidence demonstrating that models based on skill-challenge balance can effectively predict engagement levels, potentially outperforming more complex multimodal approaches

4. **Real-time Measurement Framework**: A new methodology for measuring skill and challenge directly from gameplay telemetry data, enabling non-intrusive, real-time engagement estimation in complex gaming environments, with demonstrated generalizability across different game genres beyond the primary implementation context

5. **Hybrid Technical Architecture**: A novel combination of Graph Convolutional Networks for modeling player interactions with Transformer networks for temporal sequence processing, demonstrating significant performance improvements (73% vs. 67% accuracy, 0.83 vs. 0.65 ROC-AUC)

over Transformer-only approaches by effectively capturing player spatial relationships in diverse gaming scenarios

These contributions address significant gaps in the literature and offer practical solutions for both researchers and industry practitioners interested in measuring and enhancing player engagement in modern video games.

1.7 Practical Applications

The real-time engagement metrics provided by the framework developed in this thesis offer several practical applications for game developers and researchers:

- **Dynamic Difficulty Adjustment**: Games can automatically modify challenge levels based on detected engagement states, preventing player frustration or boredom [8], [9]
- **Targeted Content Delivery**: Developers can introduce new gameplay elements or narrative sequences precisely when engagement begins to decline [10]
- **Personalized Matchmaking**: Matchmaking systems can maintain optimal skill-challenge balances across different player segments [11], [12]
- **Intelligent Resource Allocation**: Cloud gaming environments can dynamically allocate bandwidth and processing resources to maintain quality during critical engagement periods [13]
- **Game Design Optimization**: Developers can identify which specific game elements consistently drive or diminish engagement, informing future design decisions

These applications highlight the practical value of the research beyond academic understanding, demonstrating how real-time engagement measurement can enhance player experiences and potentially improve commercial outcomes in the gaming industry.

In summary, this thesis addresses the challenge of real-time, non-intrusive player engagement measurement in video games, progressing from theoretical exploration to practical implementation. By developing and validating a telemetry-based framework grounded in Flow Theory, the research offers both academic insights and practical tools for understanding and enhancing player experiences in modern gaming environments.

1.8 Results Dissemination

1.8.1 Publications

The key findings and methodologies developed in this thesis have been disseminated through the following peer-reviewed publications:

1. Ammar Rashed, Shervin Shirmohammadi, and Mohamed Hefeeda, "Real-Time Player Engagement Measurement Using Non-Intrusive Game Telemetry", *IEEE Open Journal of Instrumentation and Measurement*, Volume 4, 2025, Article Sequence Number 2500116, 16 pages.

2. Ammar Rashed, Shervin Shirmohammadi, and Mohamed Hefeeda, "Descriptor: Multimodal Dataset for Player Engagement Analysis in Video Games (MultiPENG)", *IEEE Data Descriptions*, Volume 2, 2025, pp. 17-25.

3. Ammar Rashed, Shervin Shirmohammadi, Ihab Amer, and Mohamed Hefeeda, "A Review of Player Engagement Estimation in Video Games: Challenges and Opportunities", *ACM Transactions on Multimedia Computing, Communications and Applications*, Vol. 21, Issue 7, July 2025, Article No.: 192, 33 pages.

4. A. Rashed, S. Shirmohammadi, and M. Hefeeda, "Real-time prediction of player engagement from multimodal data," *IEEE Transactions on Multimedia*, revised version in review.

1.8.2 Data Sharing

To promote research transparency and enable further work in this field, the following datasets have been made publicly available:

1. PUBG Telemetry Real-Time Player Engagement. Kaggle, DOI: 10.34740/KAGGLE/DS/7099170

2. Multimodal Player Engagement (MultiPENG). Kaggle, DOI: 10.34740/KAGGLE/DSV/10587369

1.9 Thesis Structure

This thesis is organized as follows:

- **Chapter 2: Background and Related Work** provides a comprehensive examination of player engagement concepts, measurement modalities, ground truth collection methods, and modeling approaches. It establishes the theoretical foundation for the thesis and identifies key research gaps.

- **Chapter 3: MultiPENG–Multimodal Player Engagement Analysis in Video Games** presents an exploratory multimodal investigation of player engagement, comparing various measurement approaches under identical gaming conditions and validating the effectiveness of Flow Theory-based metrics.

- **Chapter 4: Real-Time Player Engagement Measurement Using Non-Intrusive Game Telemetry** develops a practical framework for real-time, non-intrusive engagement measurement using game telemetry data, building on insights from the MultiPENG study.

- **Chapter 5: Conclusion** summarizes the key findings, discusses limitations, and outlines directions for future research in player engagement measurement.

Through this structure, the thesis presents a cohesive narrative that progresses from theoretical understanding to practical implementation, addressing the research questions through both controlled experimentation and real-world application.

Chapter 2

Background and Related Work

This chapter makes verbatim reuse or rephrasing of the material in the following paper, with permission [15]:
Ammar Rashed, Shervin Shirmohammadi, Ihab Amer, and Mohamed Hefeeda, "A Review of Player Engagement Estimation in Video Games: Challenges and Opportunities", *ACM Transactions on Multimedia Computing, Communications and Applications*, Vol. 21, Issue 7, July 2025, Article No.: 192, 33 pages.

2.1 Introduction

The video gaming industry now generates more revenue than the music and movie industries combined [1]. As a form of entertainment, gaming companies strive to "out-fun" competitors and attract more customers, making the maximization of player engagement crucial. To achieve this, game developers must first gauge players' engagement levels across entire gameplays or specific segments. While players could be interrupted mid-game to assess their engagement level, this approach is both intrusive and unscalable. Automatic estimation, if accurate enough, would be more practical. However, as games grow more extensive and complex, the automatic estimation of player engagement becomes increasingly challenging. This chapter reviews methods for modeling and estimating player engagement automatically from multimodal data, including physiological signals such as heart rate, respiration, and skin conductance; neurological signals such as Electroencephalogram (EEG); facial signals such as facial expressions, eye data, and head movements; eye metrics such as pupil size, blink rate, and gaze movements; and gameplay features such as game telemetry, pixel data, user inputs, player skill, and game difficulty. This literature review aims to provide a comprehensive foundation for understanding the complex aspects of player engagement estimation relevant to this thesis.

The gaming industry has been proactively developing tools and strategies for engagement measurement and enhancement. Recent industry reports indicate several key trends: a shift towards immersive gaming experiences, with nearly half of the top 30 games being immersive [16]; widespread adoption of live-service models showing 50% expansion [17]; integration of AI for dynamic player interactions [18]; and implementation of cross-platform experiences to maintain consistent engagement [19]. Companies are increasingly employing data-driven personalization [20] and collaborating with content creators [21] to enhance player engagement. The

industry commonly tracks metrics such as Daily Active Users (DAU), Monthly Active Users (MAU), retention rates, session length, and player churn to quantify engagement [22], [23]. Analytics platforms offer sophisticated tools for tracking event-based player retention and custom engagement triggers [24], enabling developers to make data-driven decisions about game design and player experience. This overview of industry practices provides context for the academic approaches examined in this chapter and highlights opportunities for bridging theoretical frameworks with practical applications.

Player engagement is a multidimensional concept involving cognitive, emotional, and behavioral aspects of gameplay [2]. It interrelates with the player's motivation to start or continue playing, influenced by factors like peer pressure, promotional materials, and franchise affinity. Engagement has been studied through various social and psychological theories, each defining it in terms of aspects like skill-challenge balance, cognitive load, or game world immersion [25]. This conceptual complexity complicates establishing a standard definition and scale for engagement. This chapter focuses specifically on works that quantify, estimate, predict, or measure engagement rather than delving into its theoretical, psychological, and social aspects or qualitative analysis.

Another complexity in estimating player engagement is the variety of gaming domains and platforms. AR and VR applications emphasize immersion and presence [26], [27]. Mobile games, often played through touch events and in flexible settings, introduce unique engagement elements, such as advertisement segments, distinct from PC or console gaming [28]. Serious or educational games focus on improving retention and learning performance, offering a different perspective on engagement compared to games focused solely on fun. The scope of this review is limited to entertainment video games, with other domains reserved for potential future work.

Player engagement manifests differently across various gaming scenarios, each requiring appropriate measurement considerations. In single-player games, engagement primarily derives from narrative immersion, progression systems, and the balance between preset challenges and player skill [29]. Local multiplayer introduces social dimensions where engagement features include interpersonal dynamics and immediate social feedback [30]. Online multiplayer presents the most complex scenario, combining individual performance metrics with social engagement indicators [31]. These contextual variations highlight the challenge researchers face in developing comprehensive measurement methodologies that can address the diverse manifestations of engagement across different gaming environments.

The multidimensionality of player engagement has led to various modalities for its estimation. Physiological signals include heart rate variability, blood pressure, and electrodermal activity [5]. EEG signals have been central in many studies, where an *engagement index* is often designed based on predefined EEG frequency bands and channels [32]. While rich in data, these physiological signals are noisy due to encoding information from various body functions, making accurate engagement estimation challenging. Other research has focused on observable signals from the head, face, and eyes. Head pose and movement can indicate posture changes, like leaning forward or backward, reflecting different emotional states [33]. Facial features have been used to model engagement, either through deep

learning models [34] or high-level facial action units (FAUs) [35]. Additionally, eye-tracking studies have estimated engagement levels by analyzing gaze movement, pupil dilation, and blinking rates [36]. Extending beyond the player's body, gameplay pixel data and user inputs (e.g., keystrokes) have also been used to estimate engagement [37]. This chapter summarizes the various features used to estimate player engagement and provides a taxonomy for their modalities.

Finally, mapping these features to an explainable scale of player engagement is complex and requires accurate ground truth labels for analysis and validation. The subjective nature of player engagement complicates estimation, often relying on self-reports or observer annotations. Players might complete questionnaires about their emotional and cognitive state, such as the Game Experience Questionnaire [38], or annotate gameplay recordings with perceived engagement levels [37]. Interviews and focus groups offer additional subjective assessments. To address the complexity of subjective engagement, some researchers simplify and operationalize engagement into quantifiable concepts, such as the desire to continue playing; i.e., conation [39], or by manually selecting segments of self-recorded gameplay [40]. Game levels can also be manipulated to include engaging design elements, stimulating different engagement levels in players [41]. This chapter reviews the various approaches to measuring player engagement, discussing their advantages and limitations in the context of validating engagement estimation models.

In summary, this chapter reviews the conceptual foundations of player engagement and its applications in the video gaming industry. The main focus is on taxonomizing the various modalities for estimating player engagement, ground truth measurement methods, and modeling approaches. The chapter discusses the advantages and limitations of these methods, aiming to identify research gaps that exist in the literature. This literature review serves as a necessary contextual foundation for understanding the research problems addressed in subsequent chapters of this thesis.

Figure 2.1 shows the overall framework of player engagement estimation, which also constitutes the roadmap for this chapter. The entire figure illustrates the necessary activities for building an engagement estimation model, while the dashed box shows the activities that are also performed during inference time. First, engagement must be conceptually defined, as covered in Section 2.2, where the necessary background and technical terms are explained, and a definition of engagement that captures its common essence is proposed. Next, various predictors of engagement are examined, as covered in Section 2.3, where the predictors used in the existing literature are presented and taxonomized. This is followed by determining engagement's ground truth, covered in Section 2.4, where existing methods and their comparisons are presented. The ground truth is required for the eventual validation of engagement estimation methods, which are covered in detail in Section 2.5. Finally, the chapter concludes with a summary and identification of research gaps that this thesis aims to address.

2.2 Background and Definitions

Engagement is a multidimensional construct encompassing cognitive, emotional, and behavioral aspects of the player experience. Its broad and ambiguous nature makes it challenging to define, model, or quantify precisely. Given this complex-

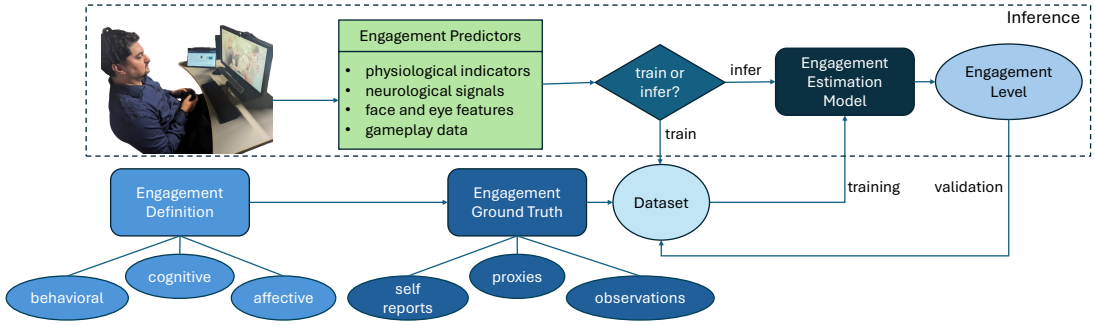


Figure 2.1: General Framework for Player Engagement Estimation in Video Games.

ity, player engagement is best studied through a combination of more concrete, complementary concepts that collectively form a mosaic-like model. This section reviews the fundamental definitions related to player engagement in games and establishes a comprehensive framework that captures the common essence of engagement across various gaming contexts.

2.2.1 Engagement as a Multidimensional Construct

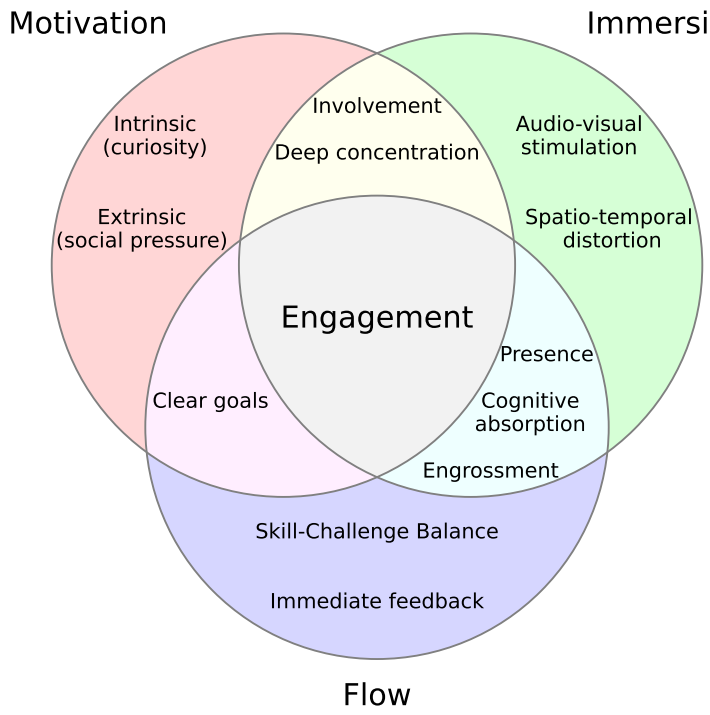


Figure 2.2: Conceptual framework of player engagement components and their interactions. The Venn diagram illustrates how Motivation (intrinsic and extrinsic factors), Immersion (audiovisual stimulation and spatiotemporal distortion), and Flow (skill-challenge balance and feedback) overlap to create the complete engagement experience, with specific constructs emerging at their intersections.

Player engagement has emerged as a crucial component in gaming research,

encompassing cognitive, emotional, and behavioral dimensions that necessitate comprehensive data collection approaches [42], [43]. It extends beyond mere interaction to include the player's immersion, motivation, and overall satisfaction with the gaming experience [44].

As illustrated in Figure 2.2, engagement can be understood as a mosaic of complementary aspects [25]. Engagement comprises three primary components that overlap and interact: motivation, immersion, and flow. Motivation can be intrinsic (e.g., curiosity) or extrinsic (e.g., social pressure) [45], [46]. Immersion encompasses audiovisual stimulation, spatiotemporal distortion, and deep concentration. Flow involves the balance between skill and challenge, along with elements like clear goals and immediate feedback. At the intersections of these components, specific constructs emerge: involvement at the motivation-immersion intersection, presence and engrossment at the immersion-flow intersection, and the central zone where all three components overlap represents complete engagement [47]–[49].

The complete engagement experience occurs at the center of the Venn diagram where motivation, immersion, and flow all intersect. This optimal state can extend beyond a single session through the concept of endurability [50], [51], where positive experiences reinforce the desire to play again. Figure 2.2 illustrates these interconnected aspects of engagement and their overlapping relationships.

2.2.2 Key Constructs in Player Engagement

Player modeling

Player modeling studies the player's experience and the dynamic phenomena of gameplay interaction [52]. It involves measuring and representing a player's skill level as a data structure [53]. While primarily based on game-player interaction dynamics, player modeling can also include static player profile information, encompassing cognitive, affective, behavioral, and demographic aspects.

Player modeling approaches can be categorized as top-down (model-based) or bottom-up (model-free). Top-down approaches use theoretical frameworks from social sciences [54], such as emotional models (e.g., valence and arousal [55]) to scale emotional states, or cognitive and behavioral models like usability theory [56] to aid game design. However, these theories often lack empirical validation in gaming contexts.

In contrast, bottom-up (model-free) approaches fit models to player data without strong a priori assumptions. These include predicting player actions, detecting behavioral patterns, or identifying player states [57], [58]. Model-free approaches bridge abstract psychological constructs with quantifiable observations, as seen in psychophysiology studies where facial expressions, head poses, and physiological signals (e.g., heart rate, EEG) are used to gauge player engagement [59], [60]. Hybrid approaches integrate model-based theories with machine learning to map player data to latent states [52].

Affect

Affect refers to aspects of gameplay that describe the player's emotional state. The Pleasure-Arousal-Dominance (PAD) model defines affect through three di-

mensions [61]: (1) *valence*, which encompasses the spectrum of emotions (e.g., pleasure to sadness); (2) *arousal*, which reflects the intensity of emotion; and (3) *dominance*, which relates to the level of control over an emotion, often referred to as *autonomy* in other contexts.

Motivation

Motivation is defined as a trait-like personal orientation toward a task [46]. It is often captured by involvement, reflecting the player's perceived relevance to a goal. The engagement process typically begins with motivation, which can be intrinsic or extrinsic [45].

Intrinsic motivation involves engaging in an activity for its inherent satisfaction, such as enjoying a game that leaves a pleasant memory and increases the desire to play again [45], [51]. Games that offer goals, such as improving skills or completing tasks, enhance intrinsic motivation. Video game designers aim to provide *novelty* to resonate with innate curiosity, a typical intrinsic motivator.

In contrast, extrinsic motivation involves engaging in an activity for a separate outcome [45]. This can range from reluctant compliance, such as studying to avoid sanctions, to self-endorsed goals, like career opportunities or social status. Ideally, games should cater to both types of motivation: offering enjoyable experiences for intrinsic motivation and rewards, incentives, or social interactions for extrinsic motivation. For instance, in multiplayer games, players might join a 'clan' and contribute regularly, even if the tasks are not inherently fun.

Immersion

Immersion is defined as a psychological state where one feels enveloped by, included in, and interacting with an environment that continuously provides stimuli and experiences [62]. Often conflated with presence, immersion relates to the player's perceived presence in a virtual environment [63]. Early research by Brown and Cairns [64] identified immersion, engagement, and engrossment as key factors of a player's experience.

Ermi and Mäyrä [65] categorized immersion into three types:

- *Sensory immersion*: Depends on the game's audiovisual execution and sensory stimulation.
- *Challenge-based immersion*: Relates to satisfying challenges and the balance between player skills and game demands.
- *Imaginative immersion*: Involves role-playing elements like story, characters, and narrative engagement.

Different game genres stimulate varying levels of these immersion types. Unlike player engagement as a whole, immersion does not directly address behavioral aspects. Jennett et al. [66] defined immersion through flow, presence, and cognitive-absorption. Flow, similar to challenge-based immersion, involves a balance between challenge and skill. Cognitive-absorption and presence correspond to sensory and imaginative immersion, respectively.

Procci and Bowers [48] summarized the construct of immersion in video games as follows:

> "Immersion is a subjective state characterized by perceiving oneself to be enveloped by, included in, and interacting with a video game that provides a continuous stream of stimuli and experiences. Immersion requires focused attention on a limited stimulus field and minimized distractions which can be promoted by the video game system itself. Immersion may be enhanced by the capability of the video game's technology to provide the player immersive cues. This includes the ability to interact with the video game through a virtual representation of the player. Interaction must seem natural with regard to the input mechanisms and the game's response to the player. Immersive cues are also strengthened by increasing the extent, fidelity and resolution of sensory information. Lacking immersive cues, involvement and individual differences may mitigate the deficit, thus helping the player to experience immersion."

As players become involved with the game, they progress from basic immersion to deeper states of cognitive absorption characterized by decreased awareness of their surroundings.

Presence

The concept of presence originally refers to the phenomenon of *externalization*, in which one's perception is referenced to an external space beyond the limits of the sensory organs [67]. In the context of video games, this space is the game environment where the player's natural perception is focused. Presence is affected by various elements in game design, such as display quality, co-player interactions, and co-playing modes [68].

Procci and Bowers [48] defined presence as:

> "Presence is a state of conviction of being located in the game environment. It is a binary experience, during which perceived self-location and perceived action possibilities are connected to the game environment, and mental capacities are bound by the game environment instead of reality."

Engrossment

Engrossment involves *emotional attachment* and *decreased perception* as described by Abbasi et al. [2], who performed interviews with participants about engaging experiences. Their research concluded that engrossment significantly affects awareness of surroundings. One interviewee noted, "I didn't notice really that it was getting darker," indicating a loss of spatiotemporal awareness. Conversely, another gamer described engrossment as choosing to focus solely on the game, stating, you have to almost seclude yourself because this is what I have to do for the next couple of hours."

In video games, graphics quality and visual realism significantly impact engrossment [69], which explains why game designers aim for visually stunning graphics to enhance player engrossment [70]. Engrossment represents a deepening of immersion, characterized by stronger emotional investment and more profound detachment from the physical environment.

Involvement

Involvement encapsulates the cognitive and motivational aspects of gameplay experiences [48]. Although distinct from immersion, involvement is reciprocally related to it as one increases the other. Generally, involvement is defined as a "person's perceived relevance of the object based on inherent needs, values, and interests" [71].

In the context of video games, Procci and Bowers [48] defined involvement as:

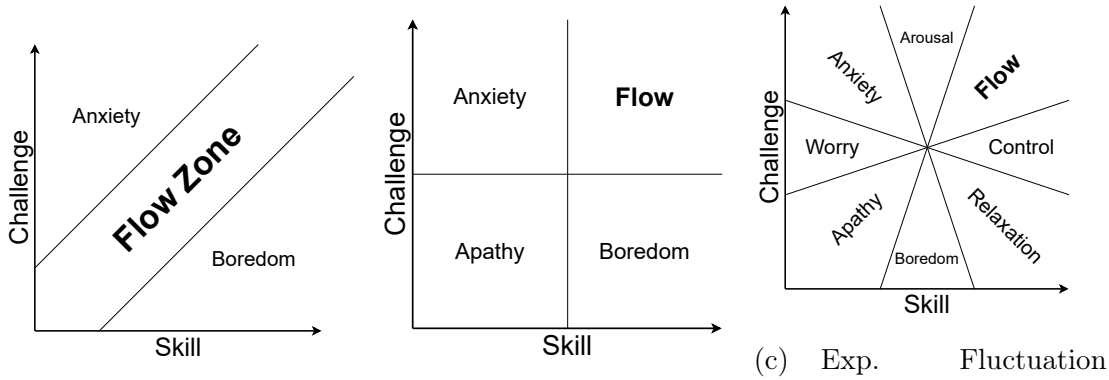
> "Involvement is a motivational factor regarding gameplay that is experienced as a sequence of focusing one's energy and attention on a coherent set of stimuli or meaningfully related set of activities or events. Involvement depends on the degree of perceived relevance that the individual attaches to the stimuli, activities, or events. Involvement is increased by playing video games that stimulate, challenge, and engage the user either cognitively, physically, or emotionally. Involvement has a reciprocal relationship with immersion, where increasing a sense of immersion similarly increases a sense of involvement, and vice-versa."

Flow

Flow Theory, introduced by Csikszentmihalyi [7], has been widely adopted in game design and engagement studies [72]. Described as the "optimal experience," flow occurs when a game presents a satisfying challenge relative to the player's skill level. Flow, akin to immersion, involves cognitive absorption where players lose awareness of their surroundings. Procci and Bowers [48] identify nine elements of flow:

1. Challenge-skill balance
2. Concentration on the task
3. Clear goals
4. Immediate feedback
5. Merging of action and awareness
6. Sense of control
7. Loss of self-consciousness
8. Time distortion
9. Autotelic experience (intrinsically rewarding)

This aligns with earlier focus on challenge/skill balance, immersion, control, and other factors [73], [74]. Flow Theory emphasizes balancing challenge and skill, with anxiety and boredom representing extremes of this balance. A well-designed game ideally transports players to their "Flow Zones," which monopolize attention, require sustained focus, reduce awareness of distractions [51], and foster pleasure and happiness [75].



(a) Basic Model. (b) Quadrant Model. (c) Exp. Fluctuation Model.

Figure 2.3: Models of Flow.

Flow's relevance to gameplay analysis lies in its focus on challenging players to reach their skill limits [46], [65]. The pinnacle of engagement is often described as flow, a state of optimal experience where players' skills are well-matched with the game's challenges.

Several models have been developed to apply Flow Theory in gaming contexts, as illustrated in Figure 2.3:

- **Flow-channel model** (Figure 2.3a): The most basic representation, where players experience high anxiety when challenges greatly exceed skills and boredom when the game is too easy. The flow zone exists as a narrow channel or corridor where challenge and skill are balanced.
- **Quadrant model** [76] (Figure 2.3b): Builds on the basic model by dividing the skill-challenge space into four quadrants. Flow occurs when both skill and challenge are high, anxiety when challenge is high but skill is low, boredom when skill is high but challenge is low, and apathy when both are low.
- **Experience Fluctuation model (EFM)** [77] (Figure 2.3c): Further refines the framework by dividing the player experience into eight distinct psychological states based on different combinations of skill and challenge levels. This model accounts for mid-levels of both dimensions, recognizing states like "worry," "arousal," "control," and "relaxation" as intermediate experiences.

While each of these models offers valuable insights into the relationship between skill, challenge, and player experience, they have limitations for real-time engagement estimation in complex gaming environments, particularly multiplayer online games where difficulty dynamically changes based on competing opponents. The models often require simplifications that may not capture the nuanced, moment-to-moment variations in player engagement that occur in real-world gaming scenarios.

Several researchers have noted that there remains significant opportunity for extending and adapting Flow Theory to better accommodate real-time engagement estimation in dynamic gaming environments [72]–[74]. The relationship

between skill and challenge continues to be recognized as a fundamental determinant of player engagement, even as new measurement approaches and technologies emerge.

Endurability

Endurability is defined as "the likelihood of remembering enjoyable situations and intending to perform them again" [51]. It involves two aspects: the Pollyanna principle, which suggests a tendency to remember pleasant experiences more than unpleasant ones [78], and *returnance*, where enjoyable experiences increase the desire to repeat them [50].

Endurability reflects how a well-designed game can boost a player's motivation to return, significantly impacting gaming behavior over time. It complements immersion and flow by focusing on the behavioral effects of enjoyable experiences, extending the engagement model beyond the immediate gameplay session.

2.2.3 Engagement Across Gaming Scenarios

Player engagement manifests differently across various gaming scenarios, each presenting unique characteristics and measurement considerations. Understanding these contextual differences is essential for developing appropriate engagement models and metrics.

In single-player games, engagement primarily derives from narrative immersion, progression systems, and the balance between preset challenges and player skill [29]. The engagement dynamics are primarily between the player and the predesigned game environment, with challenge levels typically following a carefully crafted difficulty curve.

Local multiplayer introduces social dimensions where engagement features include interpersonal dynamics and immediate social feedback [30]. The presence of other players in the same physical space adds competitive and collaborative aspects that influence engagement differently than in solo play.

Online multiplayer presents the most complex scenario, combining individual performance metrics with social engagement indicators [31]. In these environments, engagement is affected by team dynamics, competition, community interactions, and the emergent challenges created by other human players rather than pre-programmed AI.

These different gaming contexts create distinct engagement patterns. In competitive games specifically, engagement spans multiple interactive dimensions: player-versus-player (competitive combat), player-versus-environment (survival against game hazards), and cooperative team dynamics (squad coordination). The underlying Flow Theory principles regarding skill-challenge balance remain applicable across all gaming modalities, though the specific manifestations may vary considerably[79].

2.2.4 Taxonomy and Lessons Learned

The taxonomy in Table 2.1 reveals several key insights. First, engagement emerges as a dynamic process rather than a static state, incorporating multiple complementary constructs. Second, while each construct exhibits unique characteristics,

they share overlapping cognitive, affective, and behavioral dimensions, as illustrated in 2.2. Third, these constructs can be organized into a progressive chain: motivation initiates engagement, immersion deepens it, flow optimizes it, and endurability sustains it.

Consider a player's progression through this engagement chain: initially *motivated* by social recommendations or intrinsic interest, they begin playing. Audiovisual elements capture attention, evolving into deep concentration and environmental unawareness, indicating cognitive absorption. This fosters emotional attachment to game elements, leading to *immersion* and subsequently *engrossment*. The player's motivation and cognitive absorption reflect their *involvement*. With clear goals and feedback, optimal challenge-skill balance leads to *flow*, providing peak engagement. The resulting satisfaction contributes to *endurability*, reinforcing return motivation. Figure 2.4 illustrates this model.

This understanding yields two critical implications: (1) engagement cannot be reduced to a single construct but must be understood as their collective interaction, and (2) effective engagement measurement requires a multimodal approach capturing these various dimensions and their temporal progression.

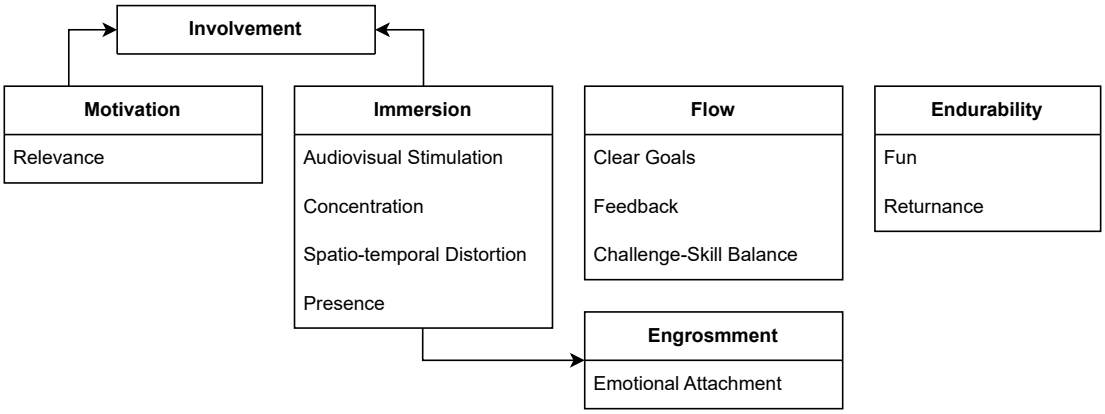


Figure 2.4: Player engagement composition as a process (left to right).

2.3 Player Engagement Predictors

In this section, we describe and categorize various signals used to quantify player engagement. As established in recent literature [42]–[44], player engagement encompasses cognitive, emotional, and behavioral dimensions that necessitate comprehensive data collection approaches. This multidimensional nature of engagement can be measured through a diverse array of modalities:

- **Physiological** indicators including heart rate, respiration, and skin conductance
- **Neurological** signals, mainly electroencephalography (EEG)
- **Facial** signals including facial expressions, eye data, and head movements
- **Eye** metrics including pupil size, blink rate, and gaze movements

Table 2.1: Qualitative comparison of Engagement-Related Constructs.

Construct	Cognitive Aspect	Affective Aspect	Behavioral Aspect
Motivation	Perceived relevance to game. Curiosity. Interest in game theme/genre.	Pleasure-driven Incentive.	Agency. Action Management. Goal-oriented.
Immersion	Audiovisual cues for sensory immersion. Spatiotemporal distortion. Concentration.	Imaginative immersion in the game world and story. Emotional attachment. Reciprocal with fun.	Volitional seclusion.
Presence	Externalization: envelopment in the game's virtual environment.	Reciprocal with emotional attachment to game world and story.	Commitment to play the game.
Engrossment	Cognitive absorption.	Emotional attachment to the game experience.	Commitment to play the game.
Involvement	Perceived relevance to player needs, values, and interests. Cognitive absorption focusing on game objectives. Reciprocal with immersion.	Reciprocal with fun.	Encapsulates motivation.
Flow	Max cognitive absorption. Challenge-skill balance.	Optimal experience. Emotional attachment. Lingering positive memories.	Clear goals. Immediate feedback. Encapsulates immersion and motivation.
Endurability	Reciprocal with immersion.	Requires strong enjoyment.	Returnance and replayability. Long commitment to the game.

- **Gameplay** features including game telemetry, pixel data, user inputs, player skill, and game difficulty

By organizing these signals into broader categories, this chapter presents a com-

prehensive taxonomy that facilitates the analysis and interpretation of engagement data within the context of game design and player experience research.

2.3.1 Modalities

Physiological Signals

Physiological signals provide insights into player engagement by recording automatic bodily responses to game stimuli. These signals can be categorized by their relation to body systems, including cardiovascular, respiratory, electrodermal activity (EDA), temperature, and muscle responses, based on traditional psychophysiology [80].

Cardiovascular Cardiovascular signals include heart rate (HR) and heart rate variability (HRV). HR can be measured using electrocardiogram (ECG) or photoplethysmography (PPG) sensors. While ECG is generally more accurate, PPG is common in wearables like smartwatches and provides comparable accuracy [81]. HRV metrics, such as the root mean square of successive differences (RMSSD) and mean standard deviation of RR intervals (SDRR), are derived from inter-beat intervals (IBI) and are used to detect arousal and stress, emotional aspects of player engagement [82], [83]. HRV is also relevant for identifying video game addiction [84]. Blood pressure, though less popular, can also be used for similar purposes [74]. The use of HRV in esports has been reviewed in [85], noting that while promising, it often lacks a solid theoretical foundation and robust methodology.

Respiration Respiration rate is measured using belts that track chest cavity expansion, with features including inspiration/expiration time, apnea, and respiration interval [86]. While some studies find no correlation between respiratory features and self-reported enjoyment [86], others use it to predict player fun levels [87]. Respiration intensity can also be measured with a digital thermometer placed under the nose, quantifying valence and arousal [88]. However, respiration features are typically used alongside other features in emotion detection, with a lack of robust ablation studies on their importance in emotion estimation models.

Electrodermal Activity (EDA) Galvanic Skin Response (GSR), or EDA, measures skin conductance, which varies with sweat gland activity controlled by the sympathetic nervous system. This response reflects emotional and cognitive processing, including reactions to threat, anticipation, and novelty [89]. EDA is useful for assessing stress, with peak height and rate serving as indicators [89]. In video games, EDA data, comprising of tonic (slowly changing) and phasic (event-related) conductance, is analyzed to gauge player engagement and cognitive load [90]. The EDA signal is sampled (e.g., at 15 Hz) and processed using techniques like Butterworth low-pass filtering to separate tonic and phasic components and detect peaks based on skin conductance response (SCR) rates and amplitudes. Increased sweating, measured by GSR, correlates with higher levels of fun [5]. While EDA and other physiological measures can predict cognitive load, they only explain part of the variance, suggesting other influencing factors [90]. EDA

features are more important in fun estimation models than features like pupil diameter, age, and perceived difficulty [87].

Temperature Skin temperature is used to assess player engagement and emotional states, with changes reflecting different emotional responses during gameplay [74]. It is measured by recording palmar temperature, which indicates autonomic or parasympathetic nervous system activation, often related to cognitive load [90]. Key features include the mean temperature and its average derivative, analyzed using statistical methods like ANOVA to detect differences under various conditions. For instance, increased game difficulty correlates with decreased temperature, suggesting a shift from engagement to boredom with higher skill levels [74]. Correlations between temperature and subjective cognitive load measures, such as frustration and NASA-TLX scale items, show medium-strength relationships [90]. However, the use of skin temperature data can be affected by confounding factors like prior activity and interactions with other physiological signals such as EDA [90]. Despite this, skin temperature remains a valuable measure for understanding player engagement and physiological responses to gameplay.

Electromyography (EMG) Electromyography (EMG) measures electrical activity produced by muscles and is valuable for assessing emotional states and cognitive processes in video games [6], [87], [90], [91]. Facial EMG sensors on muscles like the corrugator (frowning) and zygomaticus (smiling) capture changes in muscle tension related to emotional valence and mental effort [91]. EMG can detect emotional responses even without overt facial expressions [6], [91]. Other muscles, such as the biceps brachii, can be monitored for effort and cognitive load [90]. EMG data processing involves cleaning, filtering, and feature extraction to quantify muscle activation [90]. The amplitude of EMG signals correlates with cognitive load, demonstrating its relevance for understanding mental workload [90]. Additionally, EMG sensors integrated into machine learning models can predict game experience aspects like difficulty and immersion, enhancing predictive performance when combined with other modalities [6].

Neurological Signals

Electroencephalography (EEG) is a technique that measures electrical brain activity using scalp electrodes. It offers high temporal resolution, capturing rapid changes in brain activity linked to cognitive and emotional states. Recent advancements, as explored in studies like [92], [93], have enabled real-time assessment of player states and engagement during gameplay.

EEG Measurement EEG signals are processed to extract features such as power spectral density (PSD) across different frequency bands (delta, theta, alpha, beta, gamma). Each band reflects distinct neural processes [94]:

- **Delta (0-4 Hz):** Delta waves are prominent during deep sleep in adults, characterized by high amplitude and slow frequencies.
- **Theta (4-7 Hz):** Theta waves are associated with relaxed and meditative states, and their synchronization patterns modulate during changes in

affective states.

- **Alpha (8-12 Hz):** Alpha waves are observed during relaxed states and tend to diminish with cognitive exertion. They are linked to both negative and positive valence states and frontal asymmetry.
- **Beta (12-30 Hz):** Beta waves, of low amplitude, are prevalent during cognitive processes such as thinking and concentration.
- **Gamma (>30 Hz):** Gamma rhythms are associated with the binding of neural networks performing specific cognitive functions, reflect changes in affective states, and are influenced by stimuli like aversive visual cues.

Table 2.2: EEG Frequency Bands.

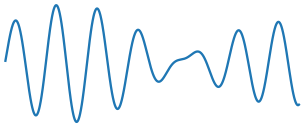
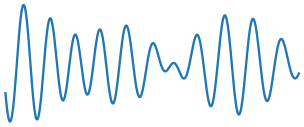
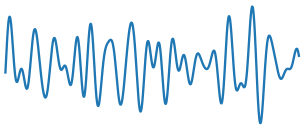
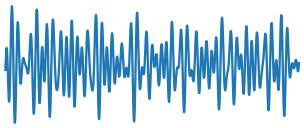
Frequency Band	**Frequency Bandwidth**	**Filtered Bandwidth**	**Neural Process**
Delta	0–4 Hz		Sleep & Dreaming
Theta	4–8 Hz		Deep Relaxing & Meditation
Alpha	8–12 Hz		Resting & Relaxing
Beta	12–30 Hz		Alert & Active Mind
Gamma	>30 Hz		Intense Focus & Problem Solving

Table 2.2 summarizes common EEG frequency bands. It is worth noting that the frequency bandwidth of these bands can differ slightly in different works.

EEG Engagement Indices Several EEG-based indices have been developed to quantify player engagement during gaming activities. These indices typically focus on specific frequency bands and their ratios, reflecting different aspects of cognitive processing and emotional arousal. The most important of these indices are:

- **Beta / (Alpha + Theta):** This index has been shown to correlate with varying levels of cognitive load and arousal during gameplay [95].

- **Frontal Theta:** This index focuses solely on the frontal theta band activity, which is associated with cognitive engagement and effortful processing. It has been shown to have relevance in distinguishing between different gaming modalities and task demands [96].
- **Frontal Theta / Parietal Alpha:** This index is particularly sensitive to changes in cognitive workload and attentional processes during gaming tasks [97].
- **Theta AF3 / Alpha P7:** Proposed in [32], this index uses the ratio of theta band activity at the AF3 electrode to alpha band activity at the P7 electrode. It aims to capture specific cognitive engagement patterns during gameplay, distinguishing between different levels of player involvement.

ML Models using EEG Using EEG data to estimate player engagement during gameplay, various classifiers have been employed, such as Support Vector Machines (SVM), Naive Bayes (NB), and k-Nearest Neighbors (kNN). For instance, [94] found that the NB classifier was most robust for identifying negative events like character deaths, while kNN, particularly using the Beta band, was better for general gameplay events, suggesting that combining classifiers can be more effective than using a single one. The work in [93] demonstrated that SVM classifiers can classify three levels of user states with reasonable accuracy, with user-dependent classification outperforming user-independent classification (66.4% vs. 50.1%). [98] showed that EEG data could classify expert and novice players with up to 98.33% accuracy using kNN. Recent studies also used genetic algorithms for feature selection and clustering, showing effectiveness in identifying EEG patterns related to player involvement [99].

These findings highlight the potential of machine learning in capturing complex EEG patterns for real-time player engagement estimation. However, model success depends on EEG data quality, feature selection, and classifier choice. Consumer-grade EEG devices may have lower spatial resolution and signal quality compared to medical-grade equipment, affecting accuracy and reliability [94]. Additionally, EEG headset comfort and placement can influence user behavior. Despite promising results, further validation against subjective measures and performance metrics is needed, and deep learning models for EEG-based engagement estimation are scarce, suggesting an area for future research.

Commercial Solutions In addition to academic research, commercial EEG solutions like EMOTIV [100] offer proprietary algorithms for measuring engagement-related metrics, reporting high accuracy in distinguishing engagement levels in controlled settings. However, as these solutions are proprietary, their methodologies are not fully available for academic validation or reproduction.

Facial

Facial signals can be crucial for understanding player engagement, offering detailed insights into emotional and cognitive states. They provide information that traditional metrics may miss [34]. Using devices like webcams, facial signals can be recorded unobtrusively and analyzed in real-time during gameplay. Advances

in computer vision have improved the accuracy of detecting subtle emotional and behavioral cues [40]. Key facial signals include:

- Facial Action Units (FAUs), which capture muscle movements linked to emotions
- Facial embeddings, used for identifying emotional patterns [40]
- Facial expressions, recognized through emotion recognition algorithms [101]
- Head movements, reflecting interest and immersion [102]

These signals provide a comprehensive toolkit for researchers and developers to enhance user experience and tailor game dynamics to player preferences.

Facial Action Units (FAUs) Facial Action Units (FAUs) capture specific facial muscle movements that indicate emotional and cognitive states, crucial for quantifying player engagement in video games. For example, FAU12 (lip corner puller) and FAU6 (cheek raiser) often signal joy or satisfaction [35]. Detected using facial landmark detection and tracking algorithms, FAUs analyze changes in facial geometry and muscle activations in real-time [40]. Studies show that FAUs effectively identify engagement levels and emotional responses during gameplay, highlighting their utility across various game genres and skill levels [34]. They enable precise monitoring of player reactions and adaptation of game dynamics, offering interpretable insights into emotional and cognitive states [103].

Facial Landmarks and Embeddings Facial landmarks are specific points on the face, such as the corners of the eyes and mouth, the nose tip, and the eyebrows, used to track facial expressions and movements in real-time [40]. Facial embeddings, on the other hand, are dense numerical representations that capture unique facial patterns for identification and emotion recognition [104]. EmoNet, for example, uses deep neural networks to estimate emotional dimensions like valence and arousal from facial images, improving accuracy under natural conditions [104].

Facial Expressions In player engagement assessment, facial expressions reveal emotions such as happiness, sadness, anger, and surprise through patterns of muscle movements [34]. While Facial Action Units (FAUs) indicate specific muscle actions, facial expressions provide a broader view of emotional states [35], [105]. This analysis helps adapt game content in real-time to enhance player immersion [106]. Facial embeddings further personalize interactions by responding to individual emotional cues, optimizing gameplay based on the player's affective state [105].

Head Pose Head pose features, which capture the orientation and positioning of a player's head during gameplay, can be used to assess player engagement. These features are extracted using computer vision techniques that analyze head angles and movements in real-time. Head pose data offers insights into a player's attention, focus, and emotional state based on their spatial interactions within the game environment [102]. Key head pose features include:

- **Yaw, Pitch, and Roll:** These are fundamental angles that describe the orientation of the head relative to a fixed reference frame. Yaw refers to rotation around the vertical axis (left-right movement), pitch around the lateral axis (up-down movement), and roll around the longitudinal axis (tilt side-to-side) [107].

- **Head Movement Dynamics:** This encompasses the speed, frequency, and smoothness of head movements during gameplay. Rapid or frequent head movements may indicate heightened engagement or reaction to game stimuli, whereas minimal movement could suggest disengagement or distraction [108].

- **Gaze Direction:** Although primarily associated with eye tracking, head pose can provide an indirect measure of where a player is looking within the game interface. Changes in head orientation relative to the screen can infer shifts in visual attention and cognitive processing [106].

- **Head Alignment:** Analyzing the alignment of the head with respect to specific game elements or events can reveal patterns of interest, such as focusing on opponents or exploring new environments [107].

Head pose features are crucial for understanding player behaviors and cognitive responses during gameplay, offering valuable insights into engagement and immersion dynamics in video games [102]. By integrating analyses of head orientation, movement dynamics, and gaze direction, game developers can tailor content in real-time to optimize challenge levels and enhance player experiences [107], [108]. These advancements underscore the growing importance of head pose recognition technologies in both research and practical applications within the gaming industry, where they contribute to refining player engagement models and improving overall game design strategies [106].

Eye

In the study of player engagement in video games, several key eye metrics have been employed to understand players' visual attention and emotional states:

- **Pupil Size:** Pupil dilation is a key indicator of arousal and cognitive load during gameplay [109]. Larger pupil size typically reflects increased arousal, which can indicate either heightened engagement or stress, depending on gameplay challenges [108]. For example, Lu et al. (2021) demonstrated that pupil size variations can reveal different levels of engagement and cognitive processing during educational game tasks, highlighting its effectiveness in assessing player involvement and cognitive effort [109].

- **Blink Rate:** Blink rate is used to infer cognitive workload and affective states like frustration and confusion during gameplay [36]. A higher blink rate often signals increased cognitive load and stress, especially in challenging game scenarios where players may experience frustration [36]. By analyzing blink rate, researchers can gain insights into how players emotionally and cognitively respond to game challenges, shedding light on their engagement levels and emotional responses [108].

- **Fixations and Saccades:** Fixations and saccades are crucial for understanding visual attention and strategy deployment in games [110]. A lower fixation/saccade ratio typically suggests more exploratory behavior or difficulty in processing game elements, which can vary with game complexity and player skill levels [36]. For instance, Burch et al. explored how these metrics reveal players' visual attention patterns across different game modes and scenarios, offering insights into effective game design and player engagement strategies [110]. Figure 2.5 illustrates the relationship between fixation and saccade.

- **Relevancy Ratio:** The relevancy ratio measures the proportion of fixations on relevant versus irrelevant game elements, offering insights into players' strategic focus and decision-making [36]. A higher relevancy ratio indicates effective attention direction towards game-relevant cues, correlating with greater engagement and task efficiency [111]. Ninaus et al. showed that this metric can distinguish between engaging and non-engaging tasks, highlighting its importance in optimizing game mechanics to sustain player interest and challenge [111].

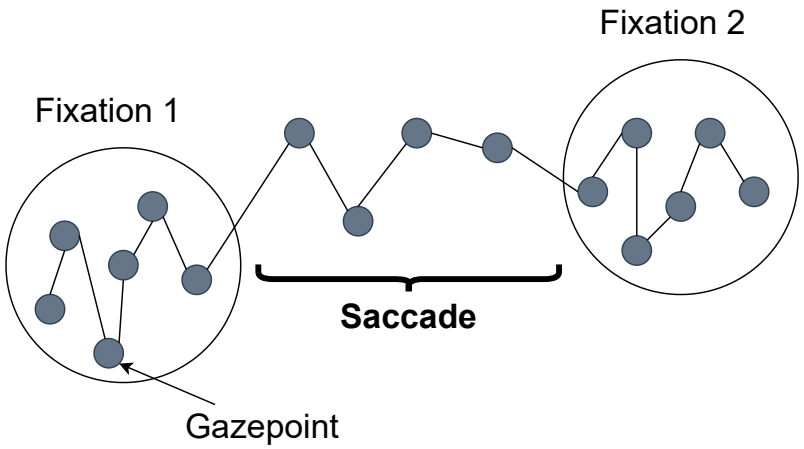


Figure 2.5: The relationship among gaze point, fixation, and saccade.

Despite their utility, eye metrics have limitations in quantifying player engagement. Their interpretation is context dependent, influenced by factors like game genre, player experience, and cultural differences, which affects their generalizability [108]. Variability in eye-tracking devices, whether high-end or webcam based, can impact the accuracy of metrics like fixation duration and saccade rates, leading to potential measurement errors and difficulties in comparing results across studies [112]. Additionally, interpreting these metrics requires considering factors such as task complexity and cognitive workload, as ambiguities can arise in dynamic and complex gaming environments [36]. Acknowledging these limitations can help refine the use of eye-tracking metrics in gaming research, enhancing our understanding of player engagement and informing game design strategies.

Gameplay

Gameplay features are crucial for quantifying player engagement, offering insights into behavior, preferences, and emotional responses. They can be categorized into:

- **User Inputs:** These include actions such as keystrokes, mouse clicks, or controller movements that reflect player decisions and interactions within the game.
- **Pixel Data:** This encompasses visual information from the game screen, including changes in the player's view, object appearances, and visual effects that contribute to the overall game experience.
- **In-Game Events (Telemetry Data):** These are data points related to game events and player actions, such as achievements, failures, or interactions with game elements, which help analyze engagement levels.
- **User Outputs:** These include user-generated content and social communications that reflect deeper engagement through creative investment and social interaction.

Utilizing these features effectively enables developers and researchers to model engagement, predict player retention, and enhance game design to keep players interested.

User Inputs User inputs refer to the interactions players have with the game through devices such as gamepads, keyboards, or mice. These inputs capture detailed actions, including button presses and joystick movements, offering insights into player behavior. They reflect immediate actions and decisions during gameplay and are crucial for understanding how players engage with the game environment and respond to challenges. For example, in a study [37] of Tom Clancy's The Division 2, detailed gamepad actions were logged and analyzed to predict player engagement. The analysis included 25 different gamepad actions and their co-occurrences, providing insights into interaction patterns. The study found that while user inputs offer detailed behavioral data, they often lack context about the outcomes of these actions within the game, making it challenging to directly infer engagement levels. Additionally, individual playing styles can introduce variability, requiring sophisticated models to extract meaningful patterns from the data.

Pixel Data Pixel data refers to frames of in-game footage that capture the visual context and the environment in which players interact. This data provides insight into the graphical and visual elements of the game, contributing to the overall gaming experience. For instance, in the said study of Tom Clancy's The Division 2, high-resolution gameplay frames were analyzed to predict long-term player engagement. The visual context from these frames was essential for understanding player behavior and actions within the game [113], [114]. However, processing pixel data is computationally intensive and requires substantial storage capacity. Additionally, while pixel data provides detailed visual information, it

may not fully capture players' emotional states or the underlying reasons for their actions. Therefore, combining pixel data with other data types, such as user inputs or physiological signals, is often necessary for a comprehensive understanding of player engagement [113], [114].

In-Game Events In-game events, or game telemetry data, offer detailed logs of player activities, progression, and interactions within the game. This data encompasses various aspects of player behavior, such as playtime, mission completion, and progression levels, providing insights into the player's journey through the game. For instance, in a study on PUBG streaming on Twitch, 40 gameplay features derived from telemetry data were used to model viewer engagement, leveraging data from hundreds of matches and over 100,000 game events to demonstrate the scalability and effectiveness of telemetry data in predicting engagement [115]. Similarly, in Tom Clancy's The Division, high-level gameplay metrics were correlated with player motivation, capturing aspects like playtime and mission completion [116]. Despite its utility, telemetry data can be coarse-grained, potentially missing the subtleties of player experience and motivation. For example, while telemetry data can indicate mission completion, it may not capture the emotional response or challenges faced by the player during that mission. Additionally, telemetry data is often game specific, which can limit its generalizability across different genres and titles [115], [116].

User Outputs User outputs provide unique insights into player engagement through user-generated content (UGC) and social interactions. UGC, including custom roles, levels, and maps, demonstrates deep engagement through creative investment in the game [117]. Social communications between players, particularly in multiplayer settings, offer rich engagement indicators through chat frequency, sentiment, and collaborative patterns [118], [119]. Viewer engagement in streaming contexts can also serve as a proxy for gameplay engagement, with chat activity correlating with engaging gameplay moments [115]. However, these measures face challenges including privacy concerns, data accessibility, and the need for sophisticated natural language processing to interpret communication content.

Industry Perspective In industry practice, gameplay features are tracked through comprehensive analytics platforms that monitor various engagement KPIs [23]. These include:

- **Session metrics:** Length and frequency of gameplay sessions, indicating immediate engagement
- **Retention metrics:** Player return rates at various intervals (1-day, 7-day, 30-day)
- **Monetization indicators:** Conversion rates, average revenue per user (ARPU), and lifetime value (LTV)
- **User acquisition metrics:** Install rates, new user acquisition, and associated costs

These industry metrics complement academic research by providing practical insights into player engagement patterns and their business impact [22], [120].

2.3.2 Engagement Across Gaming Scenarios

Player engagement manifests differently across various gaming scenarios, each requiring appropriate measurement considerations. In single-player games, engagement primarily derives from narrative immersion, progression systems, and the balance between preset challenges and player skill [29]. Local multiplayer introduces social dimensions where engagement features include interpersonal dynamics and immediate social feedback [30]. Online multiplayer presents the most complex scenario, combining individual performance metrics with social engagement indicators [31].

These diverse gaming contexts require different approaches to engagement measurement. Research by Voida et al. [30] demonstrates that in local multiplayer settings, social interactions significantly influence engagement, with indicators such as verbal communication frequency and player proximity providing valuable metrics. For online multiplayer environments, Ducheneaut et al. [31] highlight how the combination of performance metrics (e.g., combat statistics, resource management) with social interaction data creates a more comprehensive engagement profile.

The literature suggests that engagement measurement approaches should adapt to these different contexts. For example, battle royale games like PlayerUnknown's Battlegrounds (PUBG) present multiple dimensions of engagement: competitive (player-versus-player combat), survival (environmental challenges), and team coordination [115]. These games require consideration of various telemetry signals that may not be relevant in single-player narrative experiences. Similarly, engagement in role-playing games may emphasize narrative progression and character development metrics over competitive performance indicators [29].

Flow Theory principles regarding skill-challenge balance remain applicable across gaming modalities, though the specific measurement approaches require context-specific adaptation. This underscores the importance of developing flexible, non-intrusive measurement techniques, particularly for online multiplayer environments where direct observation is impractical and interruption-based measures can disrupt the gaming experience.

2.3.3 Taxonomy and Lessons Learned

The comprehensive review of engagement predictors reveals several critical insights. First, each modality offers unique perspectives: physiological signals provide objective measures of arousal and cognitive load, EEG captures fine-grained neural responses, facial signals offer unobtrusive emotional state detection, eye metrics reveal attention patterns, and gameplay features reflect behavioral engagement. Second, while each modality has distinct advantages, they also face specific limitations: physiological and EEG signals require specialized equipment, facial and eye metrics need careful contextual interpretation, and gameplay features lack emotional context and cross-game generalizability.

Most importantly, engagement prediction benefits from multimodal approaches, as each data source compensates for others' limitations. For instance, gameplay features provide behavioral context for physiological responses, while facial expressions help interpret EEG signals. Table 2.3 systematically compares these modalities, highlighting their complementary nature in capturing different engage-

ment aspects. This understanding suggests that effective engagement measurement systems should:

- Balance invasiveness with measurement accuracy
- Consider practical deployment constraints in gaming environments
- Account for individual differences and gaming contexts
- Integrate multiple modalities while respecting computational limitations
- Validate measurements against established engagement constructs

These insights inform both research directions and practical implementations in game development, emphasizing the need for context-aware, multimodal approaches to engagement estimation. The inclusion of user outputs as engagement predictors, particularly through UGC and social communications, offers promising new directions for capturing deeper, longer-term engagement patterns, though careful consideration must be given to privacy and data processing challenges.

Table 2.3: Qualitative Comparison of Player Engagement Predictors

Data Source	Features	Papers	Engagement Aspects	Practical Aspects
Physiological Signals	Cardiovascular. Respiration. EDA. Temprature. EMG.	[74] [6], [84]–[91] [121]	Game addiction. Cognitive Load. Fun. Excitement. Stress.	Measured through invasive devices. Estimatable through non-invasive widgets (e.g., smartwatches or cameras). Sensitive data raises privacy concerns. Requires stronger theoretical foundation and robust methodology.
EEG	Beta / (α + θ) Frontal θ Frontalθ / Parietal α θ AF3 / α P7	[32] [92]–[99]	Cognitive load. Arousal. Attention Patterns. Involvement.	Requires special invasive devices. Very sensitive to noise. Processing signals from several channels is computationally expensive. Generalizability of engagement indices requires further research.
Face	FAUs. Facial Landmarks. Head Pose.	[34], [35] [40] [101], [102] [104]–[108]	Emotional and cognitive states.	Accessible via webcams. Widely shared by game streamers. Pre-trained models for feature extraction are common.
Eye	Pupil Size. Blink Rate. Fixations. Saccades. Relevancy Ratio.	[36] [108]–[112]	Cognitive load. Stress. Attention Patterns.	Measured with special eye-trackers. Partially estimatable from face. Sensitive to specific game contexts and individual player characteristics.
Gameplay	User Inputs. Pixel Data. In-Game Events. User Outputs.	[37] [113]–[119]	Interaction patterns. Visual Stimuli. Motivation. Creative Investment. Social Engagement.	Noise due to individual playing styles. Cannot capture emotional responses. Cannot directly be generalized between different games or genres. Privacy concerns & data accessibility. May require language processing.

2.4 Player Engagement Ground Truth

To validate an estimated player engagement level, we must first establish reliable ground truth measurements. This section explores various approaches for establishing player engagement ground truth and validating estimation methods. We discuss several validation methodologies, including self-report methods such as questionnaires and user annotations of gameplay recordings, expert evaluation methods including observations and interviews, and heuristic and proxy methods involving concepts like conation and specially designed game levels. Each subsection examines these approaches, highlighting their application and significance in assessing player engagement in gaming contexts.

2.4.1 Approaches

Self-Report Methods

Questionnaires (QREs) Player engagement in video games is often assessed using various questionnaires designed to capture different facets of the player experience. The Game Engagement Questionnaire (GEQ) [122] was developed to provide a reliable measure specifically tailored to engagement during video game play. Validated through Rasch analysis, the GEQ is confirmed for its reliability, functional structure, and dimensionality, making it suitable for assessing engagement in gaming contexts.

While the GEQ focuses on absorption, flow, presence, and immersion, other prominent questionnaires address different aspects of player engagement. The Immersive Experience Questionnaire (IEQ) [123] measures cognitive and emotional involvement, real-world dissociation, challenge, and control, while the Player Experience of Need Satisfaction (PENS) evaluates competence, autonomy, relatedness, control, and presence/immersion. A summary of other questionnaires measuring player engagement is provided in Table 2.4, which is taken from Nordin et al. [124] with permission.

Ubisoft's Perceived Experience Questionnaire (PEQ) and the Positive and Negative Affect Schedule (PANAS) are additional tools for assessing player engagement and emotional states in gaming contexts. PEQ captures perceptions of competence, autonomy, relatedness, and presence, as used in Tom Clancy's The Division to understand motivational factors driving engagement [116]. In contrast, PANAS measures affective states like positive and negative emotions experienced during gameplay, useful in predicting engagement levels, particularly in older adults playing mobile games [60]. It was found that higher game performance, prior mobile game experience, and positive affect were significant predictors of engagement. Environmental disturbances negatively impacted engagement, while the number of gameplay sessions and game type did not significantly affect engagement levels. Interestingly, participants with dementia showed increasing engagement over time, and engagement was higher with Word-Search and Mahjong compared to Bejeweled.

Despite their utility, these instruments face challenges in universality across diverse game genres. For example, some GEQ items like "feeling scared" may not apply universally, and PENS items querying relationships with other players are irrelevant in single-player games, leading to confusion and potentially skewed re-

Table 2.4: Questionnaires measuring player engagement in video games[124].

Questionnaire	Components
Flow Questionnaire[7]	Clear goals High concentration Reduced self-consciousness Distorted sense of time Direct and immediate feedback Balance between ability level and challenge A sense of personal control Intrinsically rewarding activity
Presence Questionnaire[62]	Control factor Sensory factor Distraction Realism factor Emotional involvement
Immersive Experience Questionnaire (IEQ)[66]	Cognitive involvement Real world dissociation Challenge Control
GameFlow Questionnaire[72]	Concentration A sense of challenge Player skills Control Clear goals Feedback Social interaction Immersion
Game Engagement Questionnaire (GEQ)[122]	Absorbtion Autonomy Relatedness Presence (Immersion)
Player Experience of Needs Satisfaction (PENS)[125]	Competence Autonomy Relatedness Presence (Immersion)
Social Presence in Gaming Questionnaire (SPGQ)[126]	Psychological involvement (empathy) Psychological involvement (negative feelings) Behavioural engagement

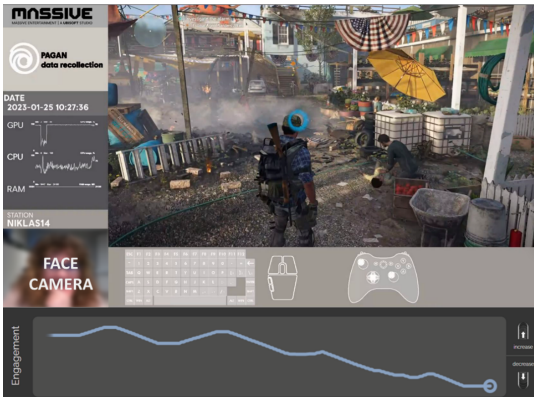
sults [123]. Moreover, the interruptive nature of questionnaires limits their ability to capture experience fluctuations throughout a gaming session, as they are typically administered post session, making them unsuitable for real-time engagement measurement.

Continuous Annotations While questionnaires are useful for gauging general player dispositions, continuous annotations offer the significant advantage of capturing emotional fluctuations during gameplay, providing a dynamic and detailed understanding of player experience. For instance, the RankTrace annotation tool [127] from the PAGAN platform allows players to continuously annotate their engagement while watching gameplay videos, facilitating the capture of nuanced changes over time [37]. This unbounded annotation approach preserves the relative relationships between data points, making subjective experiences easier to interpret [128]. Moreover, continuous annotation reduces guesswork in absolute scales, ensuring a more accurate representation of players' emotional states.

Several studies have effectively employed annotation tools to collect self-reported engagement data. In a study [37] on "Tom Clancy's The Division 2," partici-

pants used the RankTrace tool for continuous engagement annotation, enabling researchers to predict long-term engagement accurately by integrating gameplay footage and controller input. Figure 2.6 presents examples of data collection tools used for continuous annotation. Annotations are typically conducted using a wheel interface, like the Griffin PowerMate or Gazepoint Biometrics Kit, facilitating continuous and analog self-reported engagement labels. Similarly, the AGAIN dataset employed RankTrace to gather arousal labels across nine games, collecting over 37 hours of annotated gameplay videos, thereby demonstrating the effectiveness of continuous annotation in capturing diverse emotional responses across various genres [128]. These examples underscore the utility of annotation tools in providing detailed, contextually rich engagement data for advanced player modeling and affective computing research.

Despite their advantages, continuous annotation methods face limitations. A key challenge is recall bias; when players annotate engagement after gameplay, the time gap can lead to inaccuracies as they may struggle to remember and report their feelings from earlier moments [4]. Additionally, while continuous methods like RankTrace reduce some guesswork, they still rely on players' subjective interpretation of their engagement, which can introduce variability in the data. Lastly, the requirement for players to watch their gameplay videos for annotation purposes can be time-consuming, potentially affecting data reliability due to annotation fatigue, even if measures like speeding up video playback are employed [37], [128].



(a) Data Collection Snapshot via PAGAN Annotation, reused with permission from [37].

(b) Griffin PowerMate

(c) Gazepoint Biometric

Figure 2.6: Continuous Annotation Example Tools and Interface

Observational Methods

Observational methods, such as third-person expert annotations and interviews, have also been used for quantifying player engagement in video games. These methods offer insights often missed by automated systems or self-report measures. Third-person annotations add objectivity, as expert annotators — usually experienced players or researchers — apply consistent knowledge to evaluate engagement, reducing the bias from individual players. Moreover, interviews capture

nuanced experiences and motivations that players may struggle to express quantitatively, enriching the data collected.

Self-report methods assume players have high self awareness to accurately report their emotions. However, many studies employ a *third-person* annotation protocol, where expert teams assess player engagement based on observation or interviews. This is common in affect corpus compilation. For instance, in the RECOLA [129] and SEWA [130] databases, experts annotate socio-affective data from participants involved in collaborative tasks or discussions.

Several studies exemplify these methods. In one, tension was annotated by two expert Hearthstone players using the PAGAN continuous annotation tool. The annotators assessed tension during competitive matches, considering gameplay features and players' facial reactions. This approach enabled highly accurate tension prediction, underscoring the value of expert annotations in understanding emotional responses in high-stakes gaming environments [131]. Similarly, interviews with highly engaged video game players in another study uncovered key motivational factors like socialization, challenge, and positive affect, crucial for understanding continued engagement [132].

However, these methods can suffer from variability in annotations due to differences in expertise and interpretation among annotators. In the Hearthstone study, despite the annotators' expertise, inconsistencies in perceived tension may have occurred due to their subjective understanding of the game's dynamics [131]. Additionally, [133] highlighted challenges in achieving consistency in participant responses during interviews. The open-ended nature of interviews can lead to varied interpretations of questions, and a language barrier with reaction cards in their engagement mapping method emphasized the need for careful design and testing of such tools [133].

Game-Design Enforced Engagement Level

Other approaches involve simulating various engagement levels through deliberate game-level design elements [41]. These controlled settings intentionally craft environments within games that range from highly stimulating to deliberately mundane or challenging, often used in adaptive game design [134]. This method manipulates player engagement by altering factors like visual complexity, difficulty, or narrative progression. However, a limitation of this approach is the assumption that players universally perceive each game-level design with the intended level of engagement. Individual player preferences, cognitive styles, and prior gaming experiences can significantly influence how these design elements are perceived, potentially leading to varied engagement levels that may not align perfectly with the designer's intentions. Thus, while valuable for exploring the impact of design choices on engagement, these methods must consider the diverse and subjective nature of player responses to game environments.

Proxy Methods

Proxy methods involve various approaches to understanding and quantifying player engagement indirectly. These methods use game design elements and observable behaviors as proxies for deeper psychological states, offering insights into engagement dynamics without relying solely on self-reports or observational methods.

Their non-intrusive nature allows for broad applicability and easy implementation across diverse gaming contexts.

A prominent proxy method in the literature is *continuation desire*, or *conation*. This method measures a player's intrinsic motivation to persist in gameplay, seen as a fundamental indicator of engagement [34]. By assessing players' willingness to continue across different stages or levels, researchers quantify engagement through observable behaviors rather than explicit self reports. Studies show that continuation desire strongly correlates with players' emotional states and interaction with game challenges [135]. Conation is linked to the *happy-gets-happier* effect, where initially engaged players become more engaged over time as they experience success and mastery within the game environment [39]. In games like Candy Crush Saga, engagement follows a power-law behavior, with players investing more time as they progress, illustrating the compounding effect of positive engagement [39]. This approach offers a quantitative measure of engagement and insights into how game progression and difficulty influence player motivation and commitment.

However, conation's narrow focus on goal-directed behavior and persistence may overlook other critical components of engagement, such as emotional immersion, cognitive absorption, or social interaction within the game. This limitation may not fully capture instances where players are emotionally invested but not necessarily driven by specific objectives, or where engagement fluctuates due to complex emotional responses during gameplay.

By analyzing how players interact during different gameplay phases — such as intense combat versus downtime or exploration — researchers can infer engagement levels from behavioral cues and game state transitions. The FaceEngage annotation protocol [40] utilizes the picture-in-picture (PIP) format of gameplay videos, categorizing game status (active or transitional) in the main screen and user play status (busy or idle) in the inset window. High engagement is inferred from active game status with busy user play, while low engagement is associated with transitional game status and idle user play. Uncertain cases are excluded to maintain data quality and reduce bias.

A limitation of this annotation method is that it may conflate player-audience engagement with player-game engagement, potentially skewing results. The protocol relies primarily on visible game states and player actions, without accounting for cognitive states or emotional responses that aren't immediately apparent. For instance, a player might appear engaged due to active gameplay but could be emotionally disconnected or distracted. This surface-level assessment may not fully represent the complex, multifaceted nature of player engagement, particularly in diverse gaming scenarios or when players interact with their audience.

2.4.2 Multimodal Datasets for Player Engagement

Comprehensive player engagement analysis benefits from multimodal datasets that combine various measurement approaches. Most existing datasets have focused on single modalities or limited combinations. The FaceEngage dataset [40] demonstrated the value of facial expressions for engagement detection but relied solely on game status for ground truth labels. The EngageMon dataset [28] showed the potential of sensor fusion in mobile gaming contexts but was limited in scope. More recently, player engagement estimation has evolved with datasets like the

Division 2 corpus [37], which demonstrated 72% accuracy in predicting long-term engagement using game footage and controller inputs, and GameVibe [136], which provided annotated gameplay sessions across 30 diverse games with third-person affect traces.

While datasets like RECOLA and SEWA [129], [130] established protocols for synchronized multimodal collection, they were not gaming-specific. The Experience Sampling Method (ESM) [137] has emerged as an important approach for collecting ground truth engagement levels during natural game pauses, minimizing both gameplay disruption and recall bias.

Current multimodal approaches in the literature typically combine physiological signals like heart rate data [49], neurological measurements through EEG recordings [44], comprehensive eye-tracking metrics [48], user input patterns [87], webcam footage for facial analysis [40], and gameplay frames [115]. However, as noted by O'Regan et al. [42] and Poeller et al. [43], there remains a gap in the literature for comprehensive gaming datasets that synchronize all these modalities together. Particularly lacking are datasets that capture engagement across varied gameplay scenarios and difficulty levels to enable analysis through Flow Theory [7], where player skill and challenge interact as key determinants of engagement [72], [76].

2.4.3 Comparison and Lessons Learned

The comprehensive review of engagement ground truth methods reveals that each approach presents distinct trade-offs: questionnaires provide thorough assessment but interrupt gameplay, continuous annotations enable temporal tracking but suffer from recall bias, observational methods offer expert insights but introduce subjectivity, game-design approaches ensure controlled environments but may not generalize, and proxy methods enable unobtrusive measurement but can oversimplify engagement.

Our analysis through four key criteria (Table 2.5) - real-time capability, objectivity, comprehensiveness, and bias minimization - demonstrates that no single method provides a complete solution. Continuous annotation methods achieve the best balance, satisfying three criteria, while interviews face the most limitations. These findings yield crucial implications for engagement validation:

- Real-time measurement often conflicts with comprehensive assessment
- Objective scales improve reliability but may oversimplify engagement complexity
- Effective validation requires combining complementary methods
- Validation protocols must actively address temporal aspects and bias mitigation

These insights emphasize the need for multi-method approaches in both research methodology and practical game development, carefully balancing measurement accuracy with implementation feasibility. Several recent studies have employed the Experience Sampling Method (ESM) [137] to collect ground truth

engagement levels during natural game pauses, minimizing both gameplay disruption and recall bias. This approach allows researchers to complement self-reported engagement with third-party annotations from trained judges, providing both subjective and objective perspectives on engagement indicators across varied gameplay scenarios and difficulty levels.

Table 2.5: Comparison of player engagement validation methods.

Method	Papers	Real-Time	Objective	Comprehensive	Unbiased
QREs	[7], [60], [62], [66], [72], [116], [122]–[126]	✗	✓	✓	✗
Cont. Annot.	[37], [128]–[131]	✓	✓	✓	✗
Interviews	[132], [133]	✗	✗	✓	✗
Level Design	[41], [134]	✗	✓	✗	✓
Proxy	[34], [39], [40], [135]	✓	✓	✗	✓

2.5 Player Engagement Estimation Models

Now that we have reviewed the required concepts and their related literature about player engagement, including multimodal features and validation methods, we offer in this section a review of the state-of-the-art methods and models for the automatic estimation of player engagement. We focus on video games, highlighting trends in the literature. Approaches vary, employing combinations of signals and techniques such as hierarchical Bayesian models, convolutional neural networks (CNNs) with long short-term memory (LSTM) networks, and other innovative methods. By examining these methodologies, we aim to elucidate their advantages, limitations, and unique contributions to understanding player engagement dynamics. We divide the existing estimation methods into 3 groups: face-based, physiological and behavioral-based, and multimodal techniques, described next.

2.5.1 Face-based Models

These models aim to infer player engagement levels from facial expressions and, in some cases, incorporate audio data for improved accuracy. All recent models employ various machine learning techniques. Each approach presents unique strengths and limitations that highlight gaps in the current literature, discussed next.

FaceEngage Dataset

Chen et al. [40] introduced the FaceEngage dataset to address non-intrusive engagement estimation in gaming using user-contributed gameplay videos. They proposed two methods: (1) Feature Extraction & Traditional ML, which uses fixed-length feature vectors from facial motion data with traditional classifiers, and (2) Deep Learning Approach, employing a pre-trained VGG-Face CNN for face embedding extraction, followed by an encoder-decoder RNN with an attention mechanism for temporal processing. The deep learning approach significantly outperformed the traditional method, achieving an accuracy of 83.8%. This study

demonstrates the potential of using facial expressions for non-intrusive engagement estimation in gaming contexts. However, it faces potential overfitting due to limited training data, especially for the deep learning model. Furthermore, it uses a simplistic annotation approach, explained in section 2.4.1, which may not capture the full complexity of engagement, and the study does not consider the impact of individual differences in facial expressiveness or cultural variations in expression.

Facial Expressions for Conation Prediction

Rae et al. [34] developed a system to measure emotions and continuation desire (conation) during gameplay using two machine learning algorithms: an emotion recognition system trained on two million images, and a continuation desire predictor using LSTM networks. The integrated system processes facial expressions in real-time, making it suitable for live gaming environments. Validation on a different game showed 78.48% accuracy in predicting continuation desire, demonstrating some generalizability. However, the system's reliance on conation as a measure of engagement may not fully capture the nuances of emotional immersion or cognitive absorption in gameplay experiences. The study observed less pronounced emotional expressions in single-player games, which may limit the system's effectiveness in such contexts. Additionally, the approach does not account for engagement that may persist even when a player doesn't explicitly desire to continue playing.

These face-based engagement estimation models show promise in non-intrusive player engagement assessment, but they also highlight significant aspects in the current literature:

1. **Non-intrusive measurement**: Both studies demonstrate the potential of using facial expressions for unobtrusive engagement estimation in gaming contexts.

2. **Deep learning superiority:** The FaceEngage study shows that deep learning approaches significantly outperform traditional methods in this domain.

3. **Real-time applicability**: Rae et al.'s system [34] processes facial expressions in real-time, making it suitable for live gaming environments.

4. **Limited engagement definitions**: Both studies focus on narrow aspects of engagement without considering its multifaceted nature.

5. **Contextual challenges**: The models often struggle to account for varying gameplay contexts, such as single-player vs. multiplayer scenarios.

6. **Ground truth reliability**: The simplistic annotation approaches may not capture the complexity of engagement.

Future research should address these limitations by developing more comprehensive engagement models that account for its multidimensional nature, consider diverse gaming contexts, and employ more sophisticated multimodal integration techniques. Additionally, establishing more reliable methods for ground truth annotation and exploring ways to personalize engagement estimation models could significantly advance the field.

2.5.2 Physiological and Behavioral-based Models

Recent research has explored the use of physiological and behavioral data to estimate player engagement and enjoyment in gaming contexts. These approaches aim to provide more objective measures of engagement compared to traditional self-report methods, potentially offering real-time insights into player experiences.

Predicting Fun from Physiological and Behavioural Data

Fortin et al. [87] developed real-time engagement models using diverse data sources and machine learning techniques, aiming to improve adaptive gaming systems by accurately assessing player enjoyment. The study collected comprehensive data from 218 participants playing Ubisoft's Assassin's Creed games. The researchers gathered an extensive array of data, including physiological measures (ECG, respiratory activity, EDA, EMG, eye movements), questionnaires, game events, and continuous fun ratings. They extracted 244 features from this data, divided into time-dependent and time-independent categories. Initially, regression models struggled to predict fun ratings due to noisy labels. The researchers then shifted to classification methods, categorizing fun into low, neutral, and high states. Among various classifiers tested, the XGBoost classifier outperformed others, achieving the highest F1 score. This classifier highlighted the importance of features from respiratory activity, ECG, eye tracking, questionnaires, and EMG. The study found that the XGBoost classifier was particularly effective at predicting high fun states but less so for neutral fun states. This may be due to participants generally reporting high levels of fun during gameplay. An alternative ranking method, which classified changes in fun levels instead of predicting absolute values, did not improve accuracy and increased label noise. Despite its comprehensive approach, the study's high proportion of positive fun ratings may have biased results, and the imbalanced gender distribution (184 males vs. 9 females) could affect feature importance and generalizability.

Bayesian Hierarchical Models for Motivation Prediction

Sawyer et al. [138] explored the use of Bayesian hierarchical models to predict player motivation from in-game actions in educational interactive narrative games. Using the Crystal Island game, they modeled engagement across multiple contexts, capturing both general trends and specific differences between player groups. The Bayesian hierarchical linear model, trained using Markov Chain Monte Carlo (MCMC) sampling, outperformed pooled and context-specific models in predicting player motivation. This approach proved particularly effective in diverse classroom settings where context significantly affects engagement. The model's ability to provide posterior distributions offered valuable insights into how various game features influence engagement across different demographics and environments. However, data variability and availability limited the model's generalizability to broader populations or new game versions.

These above two physiological and behavioral data-based engagement estimation models demonstrate promising approaches to understanding player engagement, highlighting several key points and challenges:

- **Multimodal data integration**: Fortin et al.'s study [87] showcases the

potential of combining various physiological and behavioral signals for engagement estimation.

- **Context-aware modeling**: Sawyer et al.'s Bayesian hierarchical model [138] demonstrates the importance of considering different contexts and player groups in engagement modeling.
- **Temporal dynamics**: Both studies highlight the challenge of capturing rapid fluctuations in engagement during gameplay.
- **Demographic considerations**: The studies emphasize the need for more diverse and representative participant pools in future research.
- **Generalizability**: Both studies were limited to specific games or genres, raising questions about model applicability across different types of games.
- **Balancing objectivity and subjectivity**: While physiological measures offer objective data, correlating them with subjective experiences of engagement remains challenging.

Future research in this area should focus on developing more robust, generalizable models that can account for individual differences and diverse gaming contexts. Additionally, exploring ways to combine physiological and behavioral data with other engagement indicators, such as facial expressions or voice analysis, could provide a more comprehensive understanding of player engagement.

Industry Approaches

While academic research focuses on sophisticated modeling techniques, commercial platforms offer streamlined approaches to engagement estimation. For example, GameAnalytics' event-based tracking system allows developers to define custom triggers for specific player actions like completing challenging levels or making in-game purchases, enabling precise correlation with long-term engagement [24]. Their platform tracks 22 key metrics across engagement, monetization, and advertising categories, including granular measurements of session length, player count, and churn rate [23]. Similarly, SonaMine's analytics suite emphasizes metrics that directly tie to business outcomes, such as the stickiness ratio (DAU/MAU) for measuring regular player engagement, and detailed conversion tracking for monetization effectiveness [22]. These industry tools prioritize actionable insights - for instance, allowing developers to identify exactly which game levels or features correlate with higher retention rates [120]. This data-driven approach complements academic research by providing specific, measurable validations of engagement theories and highlighting concrete areas where theoretical models can improve game design and player retention.

2.5.3 Multimodal Techniques

Recent research has explored multimodal approaches beyond the traditional single-mode methods of incorporating physiological signals. In the following two such multimodal approaches are described.

Face, Pixel, and Audio

Pan et al. [139] extended the FaceEngage dataset by proposing a multimodal deep learning model that incorporates facial, pixel, and sound modalities. Their approach includes a face modality using a Multi-Task Cascaded Convolutional Neural Network and EfficientNet-B0 [140], a pixel modality processing entire video frames, and a sound modality extracting features from Mel-Frequency Cepstral coefficients. The model achieves 77.2% accuracy, demonstrating the potential of multimodal approaches. Notably, the study quantifies modality contributions, revealing that the sound modality dominates with a 92.6% contribution. However, this overwhelming contribution raises questions about the necessity and efficiency of including face and pixel modalities. Furthermore, the dataset's annotation method may conflate player-audience engagement with player-game engagement, potentially skewing results, as explained in section 2.4.1. The study also does not explore more complex fusion strategies, which might yield better performance.

Gamepad and Pixels

Pinitas et al. [37] introduced a multimodal approach to predict player engagement in Tom Clancy's The Division 2, using gameplay frames and gamepad inputs. Their novel dataset comprised annotated gameplay videos and gamepad actions from 25 participants. The methodology involved processing gameplay frames with a pre-trained ResNet18 neural network and encoding gamepad actions into frequencies and combos. Separate neural network architectures were used for each modality, with a fusion model integrating features from both sources. Evaluation using leave-2-participants-out cross-validation showed that the fusion model, combining both modalities, exhibited the highest accuracy in predicting player engagement. This validated the effectiveness of multimodal approaches for engagement prediction. A notable limitation was recall bias from the gap between gameplay engagement and annotation, potentially affecting the fidelity of engagement traces.

These multimodal techniques for engagement estimation in gaming highlight several key areas for future research:

- **Modality contributions**: Pan et al.'s study [139] quantifies the relative importance of different modalities, revealing the dominance of audio in their context. These findings raise questions about the necessity of including less contributive modalities in engagement models.

- **Multimodal Integration**: Both studies showcase the potential of combining multiple data sources for more accurate engagement prediction, but also highlight the challenges in effective integration.

- **Generalizability Challenges**: Each study faces limitations in terms of sample size, game specificity, or data collection methods, indicating a common challenge in the field.

- **Recall bias challenge**: Pinitas et al.'s study exemplifies the challenge of recall bias in establishing reliable engagement annotations, a common issue in engagement research.

2.5.4 Comparison and Lessons Learned

The review of engagement estimation models (Table 2.6) reveals distinct patterns across three modeling approaches. Face-based models achieve high accuracy (78-84%) but often rely on simplified engagement definitions. Physiological and behavioral models offer objective measurements but struggle with data imbalance and limited generalizability. Multimodal approaches show promise in combining complementary signals but face challenges in effective feature integration and annotation quality.

This systematic comparison yields several critical insights for engagement modeling:

- Model complexity often trades off with interpretability - deep learning approaches outperform traditional methods but offer less insight into feature importance
- Ground truth quality significantly impacts model performance, with simplified annotations and recall bias limiting accuracy
- Context-awareness remains challenging, with most models showing limited generalizability across different games or player populations
- Multimodal integration requires careful consideration of modality contributions, as demonstrated by the dominance of audio features in some studies

These findings suggest that future engagement modeling should prioritize robust ground truth collection, context-aware architectures, and thoughtful multimodal integration while maintaining interpretability.

Table 2.6: Summary of Player Engagement Estimation Models in Video Games

Type	Study	Features	Target	Data	Model	Performance	Limitation
Face-Based	Chen [40]	Face embed., motion	Binary Eng.	700+ Videos	VGG-Face+IRNN w/attention	83.8% Acc.	Simplistic annotation
	Rae [34]	Face embed., emotions	Conation	2.6M points	LSTM	78.48% Acc.	Limited definition
Physio. & Behav.	Fortin [87]	Physiological	3-level fun	218 part.	XGBoost	0.38 F1	Data Imbalance
	Sawyer [138]	Gameplay Features	Motivation	63 students	Bayesian Hierarchical	1.469 MSE	Limited generalizability
Multi-modal	Pan [139]	Face, pixel, audio	Binary Eng.	700+ Videos	EfficientNet + GRU, Audio CNN	77.2% Acc.	Simplistic annotation
	Pinitas [37]	Game pixels, user inputs	Binary Eng.	25 part.	ResNet18 + NN	72% Acc.	Recall bias

2.6 Summary, Conclusion, and Research Directions

The subjective, multi-dimensional nature of player engagement has led to various research strands, each addressing player engagement measurement from different perspectives. In this chapter, we reviewed concepts, predictors, validation methods, and estimation models of player engagement, aiming to highlight common challenges and uncover opportunities for more coherent, mature research directions.

Our review has identified four crucial dimensions for advancing player engagement estimation in video games:

2.6.1 Conceptual Framework of Player Engagement

Conceptually, player engagement spans cognitive, affective, and behavioral dimensions. The cognitive aspect is triggered by audiovisual stimulation, manifesting in heightened concentration and reduced spatio-temporal awareness. The *fun* experienced during gameplay fosters emotional attachment, reinforcing the desire to continue playing. However, this fun is influenced by factors like the balance between game difficulty and player skill, goal clarity, and feedback availability, all of which underpin player engagement.

Engagement both influences and is influenced by the player's interaction with the game. For instance, a player may begin a game highly engaged due to its relevance, promotional anticipation, or curiosity. A player consistently experiencing high engagement levels is more likely to continue playing and recommend the game to others.

This conceptual complexity directly impacts the measurement of player engagement, particularly in defining it. For example, continuation desire, or conation, can be measured by the number of attempts at a game level [34]. The variety in engagement questionnaires stems from disagreements on the definition of engagement, leading to various questionnaires (e.g., GEQ [122], IEQ [66]) that prompt players to report different gameplay experience aspects, each based on different engagement theories. While these approaches offer a nuanced understanding, the commonality among definitions is the agreement on the cognitive, affective, and behavioral nature of engagement. Therefore, a comprehensive model of player engagement should recognize it as a dynamic process spanning these dimensions, where constructs progress from motivation through immersion and flow to endurability.

2.6.2 Modalities and Predictors of Player Engagement

While player engagement indicators can be extracted from various modalities, their reliability and practicality vary. Physiological signals, such as cardiovascular metrics and body temperature, provide objective arousal measures but require a stronger theoretical foundation and robust methodology. Since smartwatches are primarily marketed for health monitoring, there's limited interest in using them for game research. A framework is needed to integrate physiological signals from ubiquitous wearable devices with gaming platforms seamlessly.

Any modality requiring specialized devices is generally limited to exploratory studies and less suitable for widespread consumer use. For instance, while EEG

captures neural responses and can provide valuable insights into player engagement, consumer EEG devices are far less ubiquitous than earbuds or smartwatches. However, eye-trackers embedded in increasingly common virtual reality devices offer a vital opportunity to quantify engagement in VR games, especially in areas like immersion and presence by showing attention patterns.

The relatively easy access to facial input, combined with the maturity of facial expression research, has led to a growing trend of using player facial footage to reveal emotions and quantify engagement. Although practical, this method requires further validation, particularly in its robustness across various input configurations like frame rate, resolution, and lighting conditions. Additionally, while physiological and neurological features focus directly on the player's state, facial input reflects behavioral responses, which may vary based on game genre, pregame state, and player personality. Engagement frameworks involving facial input must also address privacy concerns and, in mobile games, battery life. While promising in practicality, this approach is still in its early stages regarding performance validation.

Game telemetry, gameplay footage, and user-input data are among the most unobtrusive methods and have shown promise in quantifying player engagement [37]. For systems like cloud gaming, user inputs, and game streams are already collected. This approach requires developers to implement APIs to collect and share telemetry data with gaming systems, as seen in games like PUBG, to quantify engagement. However, the generalizability of these methods across different games and mechanics remains unproven.

Industry practices emphasize practical, scalable approaches to engagement measurement, focusing on metrics like DAU/MAU ratios, session length, and retention rates [22], [23]. While academic research explores sophisticated physiological and neurological measurements, industry solutions prioritize readily available data sources and actionable insights [120]. The industry's focus on event-based tracking and custom engagement triggers [24] suggests opportunities for academic research to develop more practical, implementation-focused engagement estimation methods.

Overall, while each measurement modality offers unique perspectives, multimodal approaches yield the most comprehensive understanding by compensating for individual limitations.

2.6.3 Establishing the Ground Truth of Engagement

Establishing the ground truth of player engagement measurement is arguably the most challenging aspect of the problem due to its subjective, multidimensional nature and continuous process. The subjective and multidimensional aspects typically require comprehensive assessments involving questionnaires. However, the continuous nature of engagement means that interrupting gameplay may interfere with the player experience. Previous approaches have circumvented this issue by using third-person observation, time-continuous annotation of recent playthroughs, or proxies such as conation, operationalized as the number of attempts.

Third-person observations are prone to bias, as different observers, regardless of experience, cannot objectively capture the player's experience, leading to

inconsistent measurements. This is especially relevant in player engagement measurement, as opposed to tasks like emotion recognition, where task-related signs (e.g., sobbing indicating sorrow) are relatively easier to observe. Post-game time-continuous annotations merge the self-reported aspects of questionnaires with the real-time nature of continuous annotations to capture experience fluctuations. However, such annotations are susceptible to recall bias due to the gap between the experience and the annotation [4].

Proxy concepts like conation, though practical, are too narrow to capture the full nuances of player engagement. They are more suited to exploratory studies with large player sets, but their conceptual scope limits their validation capacity. An optimal validation method for player engagement should be continuous and self-reported, with no interruption and no gap between experience and annotation. While this is practically infeasible, combining the aforementioned approaches helps mitigate the limitations.

2.6.4 State-of-the-Art Estimation Models

Current player engagement estimation models span a wide range of approaches, from face-based methods [34], [40], [139] to physiological and behavioral data-based models [60], [87], and advanced techniques like EEG-based indices [32], Bayesian hierarchical models [138], and multimodal analysis [37]. While these models show promise in specific contexts, they face common limitations: limited sample sizes, game-specific applicability, and challenges in capturing engagement's temporal dynamics.

Face-based methods struggle with individual expressiveness variations, physiological approaches grapple with data integration, and advanced techniques often lack real-time applicability. The current estimation models, though promising, reveal critical trade-offs between complexity and interpretability, accuracy and generalizability. Industry practices emphasize practical metrics while academic research explores more sophisticated but less scalable techniques.

2.6.5 Research Directions

Based on our comprehensive review, several research directions emerge for advancing player engagement estimation:

1. **Context-aware algorithms:** Developing more generalizable models that can function across diverse gaming contexts and player demographics.

2. **Multimodal integration:** Emphasizing approaches that combine various data sources (e.g., facial expressions, physiological signals, and gameplay data) to provide a more comprehensive understanding of engagement.

3. **Temporal dynamics:** Exploring methods to capture fine-grained temporal dynamics of engagement without disrupting gameplay.

4. **Computationally efficient methods:** Creating algorithms that process heterogeneous data streams in real-time for practical applications.

5. **Academia-industry collaboration:** Bridging the gap between sophisticated research models and actionable, scalable engagement estimation methods by facilitating access to diverse datasets, practical implementation contexts, and real-world validation environments.

6. **Adaptive systems:** Implementing systems that can adjust game difficulty or content based on real-time engagement estimates, enhancing player experience.

These research directions align with the challenges identified throughout our review and reveal two significant research gaps defining the scope of this thesis:

1. The need for comprehensive multimodal datasets with reliable engagement annotations to better understand the manifestation of engagement in players themselves (Chapter 3)

2. the need for non-intrusive, real-time engagement measurement methods that can be implemented in complex gaming environments without disrupting the player experience (Chapter 4)

The subsequent chapters of this thesis address these gaps systematically. First, we explore various physiological and behavioral modalities through a controlled experiment, investigating how engagement manifests in the player across different sensing channels. This work produces the MultiPENG dataset[141], which provides synchronized multimodal data with fine-grained engagement annotations collected strategically during natural game pauses. By analyzing the relative effectiveness of different measurement techniques, we establish an important baseline finding: the practical value of Flow Theory for engagement measurement, as evidenced by strong performance when using player skill and game challenge as predictors.

Building on these insights, we then shift our focus from measuring the manifestation of engagement in the player to examining how it manifests in gameplay itself. The MultiPENG study revealed that a skill-challenge based model derived from Flow Theory achieved remarkably high performance in predicting engagement. However, this approach relied on self-reported game familiarity as a skill proxy and preset game difficulty levels as challenge measures—methods that aren't practical or scalable for real-time complex gaming scenarios. Rather than replacing our multimodal investigation, we sought to extend its key finding into practical application.

We develop a novel framework for non-intrusive, real-time measurement of engagement in multiplayer online games, using game telemetry to indirectly measure both player skill and game challenge—the key components of Flow Theory that proved so effective in our controlled study. This approach allows us to address the limitations of survey-based methods while maintaining their accuracy, demonstrating the potential for practical engagement measurement in complex, dynamic gaming environments without requiring explicit player reporting or predetermined difficulty settings.

Through this progression, the thesis presents a comprehensive exploration of player engagement measurement, from understanding its theoretical foundations to creating practical, real-time measurement systems that can be implemented in modern gaming contexts.

Chapter 3

MultiPENG: Multimodal Player Engagement Analysis in Video Games

This chapter makes verbatim reuse or rephrasing of the material in the following paper, with permission [141]:
Ammar Rashed, Shervin Shirmohammadi, and Mohamed Hefeeda, "Descriptor: Multimodal Dataset for Player Engagement Analysis in Video Games (MultiPENG)", *IEEE Data Descriptions*, Volume 2, 2025, pp. 17-25.

The dataset used in this chapter is publicly available on Kaggle (10.34740/KAGGLE/DSV/10587369).

3.1 Introduction

Building on the theoretical understanding reviewed in the previous chapter, this chapter presents an exploratory multimodal investigation into player engagement measurement through the MultiPENG (Multimodal Player ENGagement) experiment and dataset. As identified in our research directions, comprehensive multimodal datasets with reliable engagement annotations are crucial for advancing engagement estimation techniques. The MultiPENG study addresses this gap by collecting synchronized data across multiple modalities while implementing a non-intrusive approach to obtaining ground truth engagement measurements.

Player engagement assessment presents unique challenges in real-time gaming contexts. While Chapter 2 provided a comprehensive review of player engagement theory and measurement approaches, this section focuses specifically on datasets and measurement techniques that informed the MultiPENG study design, highlighting gaps that this exploratory work aims to address.

The MultiPENG experiment was designed with three primary objectives: (1) to create a comprehensive multimodal dataset enabling direct comparison between sensing modalities for engagement measurement; (2) to establish which modalities offer the most practical and accurate engagement prediction in real-world gaming environments; and (3) to validate key theoretical models of engagement, particularly Flow Theory, through empirical data collection and analysis.

This experimental work bridges the gap between theoretical frameworks described in the literature review and the practical real-time estimation methods

presented in the subsequent chapter. By collecting rich multimodal data during actual gameplay sessions across different game genres and difficulty levels, we establish critical benchmarks for engagement measurement performance and identify which approaches warrant further development for real-world applications.

A notable contribution of this study is its innovative approach to gathering engagement ground truth. Rather than relying on post-hoc questionnaires or continuous self-annotation (which disrupts gameplay and introduces recall bias [4]), MultiPENG implements an Experience Sampling Method (ESM) [137] during natural gameplay pauses, minimizing both gameplay disruption and recall bias. This methodological approach aligns with our research direction of capturing temporal dynamics without compromising the gaming experience.

3.1.1 Existing Datasets and Their Limitations

Player engagement measurement requires both reliable features and accurate ground truth labels. Existing datasets have primarily focused on single modalities or limited combinations, creating a significant gap in comprehensive multimodal approaches.

The FaceEngage dataset [40] demonstrated the value of facial expressions for engagement detection but relied solely on game status (e.g., cutscene, mission, menu) and player status (e.g., away, passively watching, actively playing) for ground truth labels without validation against player self-reports. The EngageMon dataset [28] showed the potential of sensor fusion in mobile gaming contexts but was limited in scope.

More recent contributions include the Division 2 corpus [37], which utilized game footage and controller inputs to predict long-term engagement, and GameVibe [136], which provided annotated gameplay sessions across 30 diverse games with third-person affect traces. While datasets like RECOLA and SEWA [129], [130] established protocols for synchronized multimodal collection, they were not gaming-specific.

3.1.2 Limitations of Current Measurement Approaches

Current approaches to engagement measurement face several limitations that affect their practicality and reliability:

Self-Report Methods

Questionnaires like GEQ [3], IEQ [66], and PENS [125] typically collect data after gameplay sessions, introducing recall bias and failing to capture moment-to-moment engagement fluctuations. This makes traditional surveys unsuitable for real-time measurement.

Biometric Signals

Physiological and neurological measurements from EEG [32], eye-tracking [36], and heart rate monitoring [5] provide objective indicators but require specialized equipment that limits studies to laboratory settings. The sensitive nature of biometric data also raises privacy concerns for widespread implementation.

Game-Specific Methods

Approaches based on game telemetry [37] or persistence metrics [39] are often tied to specific games or genres, limiting their generalizability. Methods that manipulate game elements to induce engagement states [41] assume uniform perception across player populations.

Computer Vision Approaches

Webcam-based methods offer non-intrusive measurement but often rely on narrow engagement definitions. Some approaches focus primarily on emotional expressions [142], [143] or continuation desire [34] rather than engagement's full multidimensional nature.

3.1.3 Requirements for Effective Engagement Measurement

Based on these limitations and the research directions identified in Chapter 2, we established five key criteria that an effective player engagement measurement approach should satisfy:

C1. Non-interrupting: The method should not disrupt the natural flow of gameplay, addressing the temporal dynamics challenge.

C2. Player-Reported Ground Truth: Ground truth should derive from players' self-reported experiences rather than indirect proxies.

C3. Minimized Bias: The temporal gap between experience and reporting should be minimized to reduce recall bias.

C4. Generic Applicability: The approach should generalize across different game genres and platforms, supporting context-aware algorithms.

C5. Practical Implementation: The method should not require inaccessible equipment or sensitive data, enabling computationally efficient methods.

Table 3.1: Criteria Comparison of Existing Measurement Approaches

Approach	C1	C2	C3	C4	C5
Surveys & Interviews	✗	✓	✗	✓	✓
Biometric Signals	✓	✓	✓	✓	✗
Game-Specific Elements	✓	✗	✓	✗	✓
Standard Webcam Methods	✓	✗	✓	✓	✓
MultiPENG Webcam	✓	✓	✓	✓	✓

The MultiPENG dataset addresses these gaps by combining multiple modalities with strategically timed self-report measures during natural gameplay pauses. This allows for direct comparison of modalities while maintaining ecological validity. Particularly, the webcam modality in MultiPENG combines the practicality of computer vision approaches with the reliability of questionnaires through strategic positioning of assessment prompts, satisfying all five criteria as shown in Table 3.1.

The insights gained from this exploratory work—particularly the performance of engagement models based on skill-challenge balance derived from Flow Theory—directly inform the telemetry-based approach developed in the subsequent chapter. By systematically evaluating which modalities and theoretical frameworks most effectively capture player engagement, we establish a foundation for developing computationally efficient, non-intrusive measurement methods suitable for complex gaming environments.

3.2 The MultiPENG Dataset

This section details the MultiPENG dataset, including its collection methodology, quality validation, and organizational structure. The dataset was designed to capture multiple modalities of player engagement simultaneously during actual gameplay, enabling direct comparison between different sensing approaches.

3.2.1 Collection Methods and Design

Our data collection system integrates multiple specialized hardware and software components to capture diverse signals indicative of player engagement. The complete experimental setup, illustrated in Figure 3.1, consists of six key components synchronized through a central gaming PC that serves as the primary data collection and synchronization hub. The setup is designed to maintain participant comfort while ensuring reliable data collection across all modalities.

A written consent was signed and obtained from all participants and the methodology was approved by the University of Ottawa's Office of Research Ethics and Integrity, under file number H-07-23-9439.

Hardware Configuration

The primary data acquisition hardware includes a 1080p webcam mounted on the monitor for capturing facial expressions and head pose, an EPOC X EEG headset operating at 128 Hz with 14 channels for brain activity measurement, and a Gazepoint GP3 Eye Tracker positioned below the monitor sampling at 60 Hz. Participants wear a Fitbit Versa 3 smartwatch on their left wrist for heart rate monitoring, while gameplay input is captured through an Xbox USB gamepad. A separate touchscreen tablet is positioned for engagement survey responses, and an additional screen is used to monitor the experiment through a unified interface.

The experiment environment is carefully controlled to simulate a natural gaming setting while maintaining data quality. The testbed operates in a sound-isolated room with consistent lighting to minimize variations in webcam and eye-tracking data. The viewing distance and monitor angle are standardized (65cm from eye tracker, monitor tilted at 15 degrees) to maintain eye-tracking accuracy across sessions. An adjustable chair ensures participant comfort throughout the session.

Software Infrastructure

The software infrastructure consists of several specialized components operating in concert. OBS Studio captures both the webcam feed and gameplay footage,

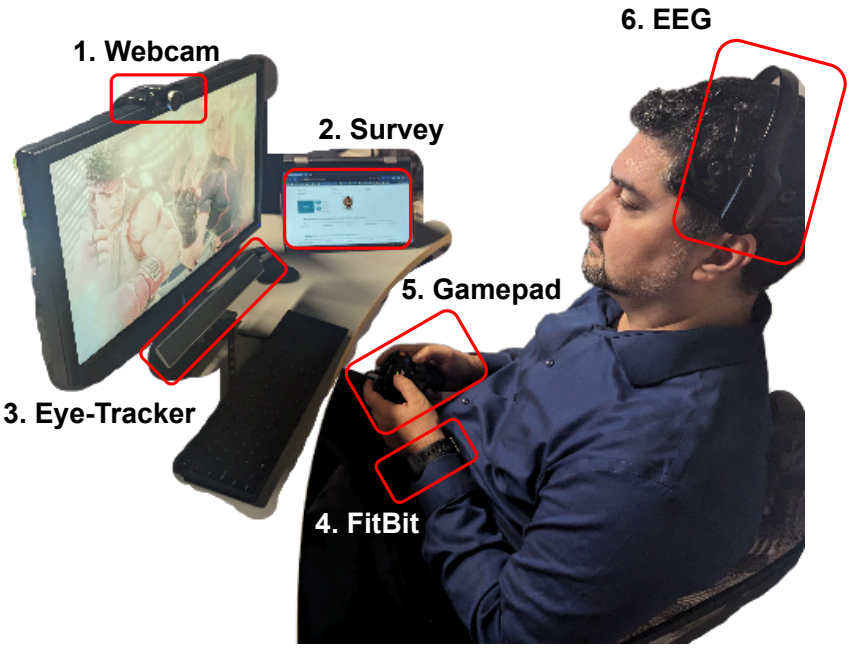


Figure 3.1: Experimental setup showing the integrated hardware components: (1) webcam, (2) survey tablet, (3) eye tracker, (4) smartwatch, (5) gamepad, and (6) EEG headset. ©ACM reused with permission from [15] / cropped and horizontally flipped from original and annotated.

with audio recorded from the webcam's built-in microphone and the game's audio output. As shown in Figure 3.2, the OBS interface serves as a unified auditing system, displaying the webcam feed, gameplay footage, system clock, EEG signal quality, and eye tracking quality metrics in a single view. This synchronized display enables real-time monitoring of data quality across all modalities and provides a comprehensive record of each session.

The collected webcam footage was processed using OpenFace [144], a comprehensive facial behavior analysis toolkit. This processing extracted detailed facial features including head pose dynamics (translation, rotation, velocity, and acceleration vectors), facial landmarks, gaze direction estimates, and 17 facial action unit (AU) intensities. These facial analysis capabilities complement the dedicated eye tracking hardware while providing additional metrics like head movement patterns that have been shown to correlate with engagement in prior work [40].

Additionally, the webcam footage was passed through EmoNet [104], to extract facial embeddings, estimate valence and arousal values, and a normalized score of 8 different emotions; neutral, happy, sad, surprise, fear, disgust, anger, and contempt.

The Gazepoint Analysis software captures comprehensive eye metrics including gaze position relative to the screen, pupil dilation, fixations, saccades, and blink rate. The Gazepoint Control software manages device calibration and maintains tracking accuracy throughout the session. The Emotiv Pro software handles real-time EEG signal acquisition and processing, providing both raw EEG signals and derived performance metrics including attention, engagement, and stress levels. Signal quality is continuously monitored through contact quality (CQ), machine

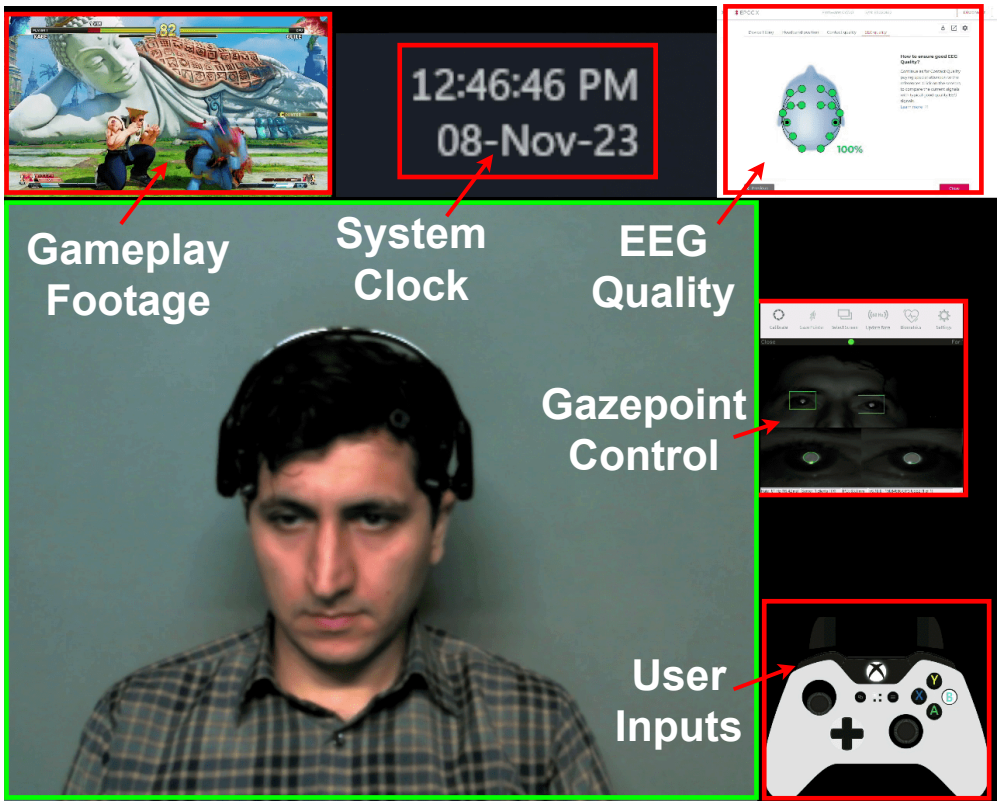


Figure 3.2: OBS Recording Scene showing the synchronized capture of webcam feed, gameplay footage, and system metrics. This unified view provides a comprehensive record of each session and enables real-time quality monitoring.

learning signal quality (SQ), and signal magnitude quality (SMQ) metrics.

Heart rate data is logged and uploaded in real-time to the Google account connected to the smartwatch at 5-second intervals, while a custom Python script handles gamepad input logging, recording timestamped controller interactions including button presses (A,B,X,Y, bumpers) and stick positions (left/right X/Y coordinates from -32767 to 32767).

Data Collection Protocol

The collection process begins with participant registration, capturing demographic information and gaming experience, particularly familiarity with soccer and 2D fighting games. The calibration sequence includes the eye tracker's standard 9-point calibration procedure using Gazepoint Control, followed by a 15-second eyes-open and eyes-closed EEG baseline recording. The Fitbit is secured snugly on the left wrist, and proper electrode contact is verified for the EEG headset using saline solution to ensure optimal signal quality.

Two popular games were strategically selected based on their representativeness, accessibility, difficulty variability, and natural session boundaries. FIFA'23 and Street Fighter V (SFV) represent two distinct and popular gaming genres (sports and fighting) with different gameplay paces and skill requirements. These games offer precise difficulty control (FIFA: 6 levels, SFV: 8 levels) allowing systematic investigation of engagement across challenge intensities. Importantly, while FIFA typically requires prior experience for meaningful play, SFV's simple

core mechanics enabled participation from players with minimal gaming background while still offering depth through advanced techniques. This complementary selection ensured our dataset captures a broader spectrum of player experiences and skill levels than would be possible with a single game type.

Both games feature natural pauses that facilitate non-disruptive survey administration—a critical design consideration for maintaining ecological validity. For FIFA'23, only participants with prior soccer gaming experience participated, playing 3-5 matches with surveys conducted after goals, at half-time, and post-match. To prevent survey fatigue while maintaining data quality, a minimum 20-second gameplay duration is enforced between consecutive surveys.

For SFV, participants undergo a 5-10 minute training phase until they report comfort with basic controls and mechanics. Each round has a maximum duration of 99 seconds, though rounds typically conclude earlier through knockouts. Participants are targeted to play three rounds at each difficulty level (low: 1-3, medium: 4-5, high: 6-8), with the actual number varying based on remaining session time and training duration. Surveys are administered between rounds, coinciding with the game's natural break points.

Survey Design

The survey application captures self-reported metrics across four key dimensions using 5-point Likert scales, as detailed in Table 3.2. The selection of these specific metrics serves two theoretical frameworks. First, while engagement serves as the primary metric, interest and excitement map to the fundamental dimensions of emotion measurement (valence and arousal), while interest connects to conation—the desire to continue playing—which is often used as an engagement proxy [34]. Stress levels relate to the flow theory of engagement [7], particularly regarding game challenge intensity. Second, these metrics mirror those reported by the EMOTIV Pro EEG software but in a gaming-specific context, enabling analysis of correlations between general EEG-based metrics and gaming-specific self-reported states. The engagement dimension ranges from Very Bored (0) to Very Engaged (4). Interest is measured from Strongly Disliked (0) to Strongly Liked (4). Stress levels span from Very Relaxed (0) to Very Stressed (4), and excitement ranges from Not Excited (0) to Extremely Excited (4).

Synchronization Implementation

The synchronization system aligns all data streams through careful clock calibration. The high-frequency data streams (EEG at 128 Hz, eye tracking at 60 Hz, and gameplay footage at 30 fps) are logged on the gaming PC with a single clock. While the smartwatch clock exhibits a 1-2 second gap with the PC clock, this discrepancy is tolerable given its 5-second sampling interval for heart rate data. The PC and survey tablet clocks are manually calibrated to the smartwatch's clock with sub-second discrepancy, ensuring temporal alignment across all data streams while accommodating the lower sampling rate of the heart rate measurements.

Table 3.2: Survey Questions and Response Options

Dimension	Question and Response Scale
Engagement	How engaged did you feel? 0. Very Bored 1. Somewhat Bored 2. Neutral 3. Somewhat Engaged 4. Very Engaged
Interest	How much did you enjoy? 0. Strongly Disliked 1. Disliked 2. Neutral 3. Liked 4. Strongly Liked
Stress	How stressed did you feel? 0. Very Relaxed 1. Relaxed 2. Somewhat Stressed 3. Stressed 4. Very Stressed
Excitement	How excited did you feel? 0. Not Excited 1. Slightly Excited 2. Moderately Excited 3. Extra Excited 4. Extremely Excited

3.2.2 Validation and Quality

To validate the quality and utility of our dataset, we present evidence supporting both our measurement framework and demonstrate the dataset's effectiveness through multiple use cases.

Quality Monitoring

We collected quality metrics for EEG, heart rate, and eye-tracking samples provided by the corresponding data collection software. The EPOC X EEG headset provides continuous signal quality metrics including contact quality (CQ), machine learning signal quality (SQ), and signal magnitude quality (SMQ). These quality metrics are aggregated into a 0-100 overall quality score indicated in column EQ.OVERALL (see Table 3.4). The Fitbit heart rate measurements include confidence levels (0-3 scale). We also use the FPOGV flag in gazepoint data as a binary quality score indicating whether there is a valid point of gaze (POG) detected. These metrics are included in the dataset, allowing researchers to establish appropriate quality thresholds for their specific analyses.

To understand the effect of different quality thresholds on the dataset size, we calculate the average quality scores per sample (i.e. an annotated game session) for each of the three modalities. Figure 3.3 shows the complementary cumulative distribution function (CCDF) of samples given a quality threshold. Interestingly, only 50% samples have an average EEG quality (EQ.OVERALL) of at least 75%. This is mostly due to sudden player movements during gameplay, which we observed to cause a short-term decline in EEG signal quality. This can be an interesting research direction to explore the relationship between sudden drops in EEG quality as a proxy for sudden movement and highly engaging gameplay moments. Similarly, the heart rate signals show relatively low confidence overall. These results emphasize the importance of post-processing physiological signals to

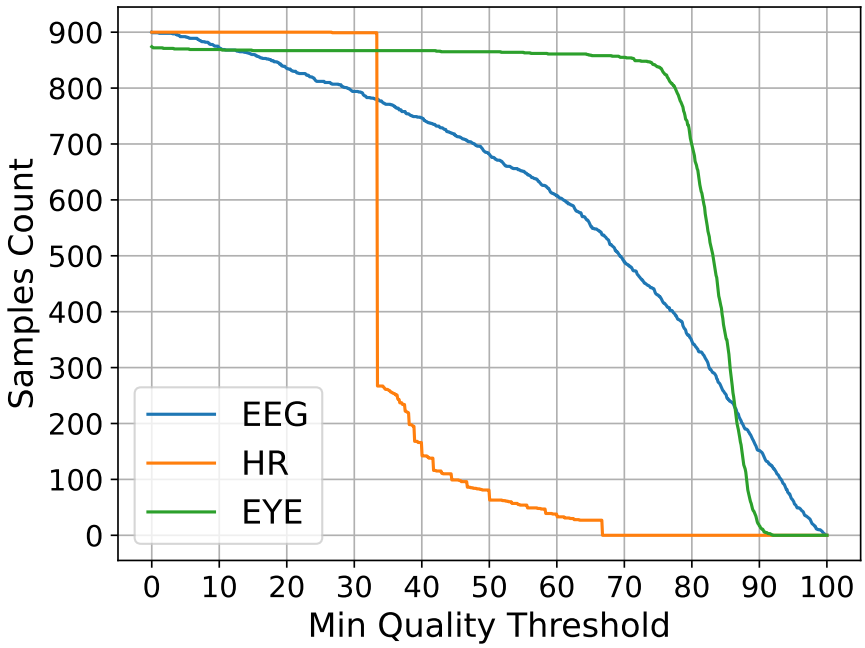


Figure 3.3: Sample survival rate vs. score threshold.

maximize their utility in modeling player engagement. Most eye-tracking samples have 75-80% valid POG. Notably, invalid FPOGV flag are often associated with blinking events, which do not necessarily indicate noisy data.

Technical Limitations and Dataset Scope

While our dataset offers rich multimodal data across different games and participants, we acknowledge certain constraints in its scope and technical implementation. The game selection, though strategically chosen to represent different genres and skill levels, is limited to two titles and may not fully represent all gaming experiences across the vast landscape of game genres. However, this limitation is partially mitigated by our diverse participant demographics, which includes individuals ranging from those with no prior gaming experience to casual players and experts in the selected games.

Prior to the main data collection, we conducted an initial trial with 4 participants to refine the experimental protocol and identify potential issues. During the full data collection, technical and procedural limitations were encountered:

- One participant (ID: 559) reported being previously diagnosed with ADHD and was using stimulant medication during the experiment. The divergence between this participant's EEG signals and the baseline signal resulted in a diminished EEG quality score, consistent with prior research on stimulant effects on EEG signals [145]

- Six participants (IDs: 872, 850, 568, 533, 297, 183) were recorded under different lighting conditions than the standard protocol

- Controller input data was not captured for ten participants (IDs: 120, 166, 462, 539, 623, 703, 754, 507, 514, 744)

- Eye tracking data was incomplete for one participant (ID: 407)

These missing data points represent a limitation that researchers should consider when analyzing the affected sessions. However, our dataset's modular structure enables researchers to selectively include participants based on available modalities for specific research questions. For instance, in our multimodal neural architecture implementation, we utilized only participants with complete data for the specific modalities being investigated. The multimodal nature of our dataset provides inherent redundancy across different data streams, potentially enabling more robust analyses even when certain modalities are unavailable for some participants. We recommend that researchers clearly document which participant subsets they use for each analysis to ensure reproducibility. Future extensions of this work could include additional game genres and address these technical challenges to build upon the foundation established by this dataset.

Validation of Engagement Dimension Selection

The theoretical framework underlying our four-dimensional survey design was validated through two complementary analyses. First, correlation analysis between metrics revealed meaningful relationships supporting our measurement approach, as shown in Figure 3.4. Engagement showed strong positive correlations with excitement ($\rho = 0.68$, $p < 0.001$) and interest ($\rho = 0.58$, $p < 0.001$), validating our connection to fundamental dimensions of emotion measurement (valence and arousal). The moderate correlation between stress and engagement ($\rho = 0.40$, $p < 0.001$) aligns with flow theory's emphasis on challenge intensity, while the weak correlation between interest and stress ($\rho = 0.11$, $p = 0.001$) confirms these capture distinct aspects of gameplay experience.

Second, comparison with EMOTIV Pro's EEG-based metrics during gaming sessions revealed important insights about engagement measurement in gaming contexts. The weak correlation between EEG-measured engagement and self-reported engagement ($\rho = 0.076$, $p < 0.05$), along with similarly weak correlations for other metrics (EEG-measured stress: $\rho = 0.089$, $p < 0.01$; interest: $\rho = 0.003$, $p = 0.92$; excitement: $\rho = -0.113$, $p < 0.001$), validates our choice of gaming-specific engagement dimensions over general EEG-based metrics. These results demonstrate that while commercial EEG systems can measure general cognitive states, gaming engagement requires domain-specific measurement approaches.

3.2.3 Records and Storage

The dataset comprises 900 micro-game sessions from 39 participants (30 male, 9 female, mean age 24.3 years), with sessions distributed across both FIFA'23 and Street Fighter V. As shown in Table 3.3, the dataset captures a wide range of engagement levels and related psychological states (interest, stress, excitement), with session durations varying significantly between games (FIFA: mean=91.5s, SD=50.3s; SFV: mean=36.7s, SD=9.8s). The comprehensive nature of this dataset, combining multiple modalities with fine-grained temporal alignment and varied gameplay scenarios, provides researchers with rich opportunities for investigating player engagement across different game genres, difficulty levels, and measurement approaches.

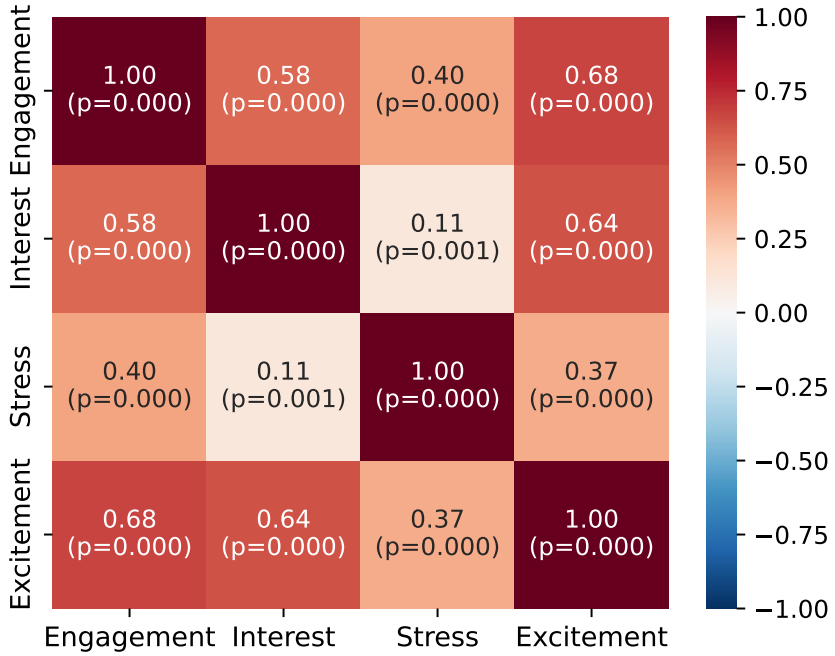


Figure 3.4: Spearman correlation matrix between engagement metrics.

The *Multimodal Player Engagement* dataset is publicly available on Kaggle[1] and is organized as follows:

```
Samples/
    pid /
        EEG/
        EYE/
        HR/
        OBS/
        OpenFace/
        EmoNet/
        XBOX/
Questionnaires/
    participants.csv
    submissions.csv
splits/
    fold_[0-6]/
        [0-6].csv
Human Panel Samples/
    Samples/
        sid.mp4
    Training/
    annotations.csv
    annotators.csv
```

The *Samples* folder contains the primary multimodal data collection organized by participant (`pid`: participant_id) and modality. The *Questionnaires* folder

[1] kaggle.com/datasets/ammarrashed23/multimodal-player-engagement

Table 3.3: Session counts and durations (minutes) per game and dimension.

Dimension	Level	FIFA		SFV		Total	
		#	Dur.	#	Dur.	#	Dur.
Engagement	0-1	14	16	88	47	102	63
	2	26	40	137	79	163	119
	3-4	97	170	538	342	635	512
Interest	0-1	6	10	90	48	96	58
	2	56	88	227	134	283	222
	3-4	75	127	446	287	521	414
Stress	0-1	53	71	282	164	335	235
	2	40	74	264	165	304	239
	3-4	44	81	217	140	261	221
Excitement	0-1	36	50	194	108	230	158
	2	56	102	294	182	350	283
	3-4	45	74	275	179	320	253
Total		137	226	763	468	**900**	**694**

contains the primary ground truth data collected during the experiment through participant surveys. The *splits* folder provides a standardized benchmark framework implementing nested stratified group cross-validation. The *Human Panel Samples* folder provides a curated subset of gameplay sessions for human evaluation of player engagement through visual cues. We explain each folder in the following.

"Samples": The Multimodal Data

Each participant's data is stored in a separate subfolder identified by their unique participant ID (`pid`), with further subfolders for each data modality (EEG, eye tracking, heart rate, OBS scene, and XBOX controller inputs). The hierarchical structure ensures clear organization of the extensive multimodal data while maintaining the relationship between different data streams for each gaming session. The naming convention of samples is `pid_sid_eng_int_str_exc`, where `sid`: submission_id, `eng`: engagement, `int`: interest, `str`: stress, `exc`: excitement. The file extension of samples is `.mp4` for OBS videos, and `.csv` for the rest.

EEG Data Files The EEG files contain measurements from an EPOC X headset with 14 channels (AF3, AF4, F3, F4, F7, F8, FC5, FC6, O1, O2, P7, P8, T7, T8). Each file begins with metadata columns: a Timestamp in datetime format (e.g., "2023-10-31 14:24:22.025-04:00"), a sample Counter (0-127, resets every second), and an Interpolated flag indicating whether the sample was received from the headset (0) or interpolated. The data is organized into several measurement categories, each sampled at different frequencies, as shown in Table 3.4.
The primary data consists of raw EEG voltages sampled at 128 Hz. For each channel, the signal quality is monitored through two metrics: Contact Quality (CQ) indicates the physical connection quality between electrodes and scalp, while EEG Quality (EQ) provides a more comprehensive signal quality assessment updated

Table 3.4: EEG Data Structure

Category	**Rate**	**Column Format**	**Values**
Raw EEG	128 Hz	EEG.{channel}	µV readings
Contact Quality	128 Hz	CQ.Overall	0-100
		CQ.{channel}	0-4
Signal Quality	2 Hz	EQ.Overall	0-100
		EQ.{channel}	0-4
Performance Metrics	0.1 Hz	PM.{metric}	Type: - IsActive (0/1) - Scaled (0-1) - Raw (unbounded) - Min, Max (bounds)
Band Powers	8 Hz	POW.{channel}	Band type: - Theta (4-8 Hz) - Alpha (8-12 Hz) - BetaL (12-16 Hz) - BetaH (16-25 Hz) - Gamma (25-45 Hz)

every 500 ms. The headset computes higher-level performance metrics at 0.1 Hz, including measures of Engagement, Excitement, Stress, Relaxation, Interest, and Focus. Each metric includes both raw algorithm outputs and normalized values. Additionally, the power in five frequency bands is computed for each channel at 8 Hz, providing insights into different aspects of brain activity during gameplay.

Eye Tracking Data Files The eye tracking data is collected at 60 Hz using a Gazepoint GP3 eye tracker. Each record contains a timestamp and a video frame counter (VID_FRAME), along with three main categories of measurements:

- The gaze data includes both filtered (FPOG) and unfiltered (BPOG) point-of-gaze coordinates. Filtered coordinates (FPOGX, FPOGY) represent fixation points with associated start time (FPOGS), duration (FPOGD), and a unique identifier (FPOGID). Unfiltered coordinates (BPOGX, BPOGY) provide raw gaze positions. Each measure includes a validity flag (FPOGV, BPOGV).

- Individual eye measurements track pupil position (LPCX/RPCX, LPCY/RPCY), diameter in both pixels (LPD/RPD) and millimeters (LPMM/RPMM), and a scale factor normalized to the calibration depth (LPS/RPS). Each measurement includes its validity flag (LPV/RPV, LPMMV/RPMMV).

- The system also tracks blink events with unique identifiers (BKID), durations (BKDUR), and frequency (BKPMIN, blinks per minute), as well as saccade characteristics including magnitude (SACCADE_MAG) and direction (SACCADE_DIR). A pixel-to-millimeter conversion scale (PIXS) is provided with its validity flag (PIXV).

Heart Rate Data Files The heart rate data is collected at 0.2 Hz (every 5 seconds) using a Fitbit Versa 3 smartwatch. Each record contains a timestamp, heart rate in beats per minute (BPM), and a confidence measure (0-3) indicating the reliability of the reading. The confidence value is determined by the smartwatch's internal algorithms based on factors like sensor contact quality and motion artifacts.

Controller Input Data Files The Xbox controller inputs are recorded asynchronously (event-driven) and consist of two types of events: analog inputs (Absolute) and button presses (Key). Each record contains a timestamp and the event details, including the specific control (Event) and its state.
Analog inputs (EventType: "Absolute") include stick positions (left_stick_x, left_stick_y, right_stick_x, right_stick_y), trigger depths (left_trigger, right_trigger), and d-pad directions (dpad_x, dpad_y). For these events, the state ranges from -32767 to 32767, representing the full range of motion.
Button events (EventType: "Key") capture binary states (0 or 1) for all controller buttons: face buttons (a_button, b_button, x_button, y_button), bumpers (left_bumper, right_bumper), stick clicks (left_stick_button, right_stick_button), and menu buttons (start_button, back_button).

"Questionnaires": The Survey Data

This folder contains the primary ground truth data collected during the experiment through participant surveys. The data is organized in two CSV files shown in Table 3.5: a participant registry capturing demographic information and gaming experience, and a comprehensive session log containing engagement metrics and contextual information. All timestamps follow the format (YYYY-MM-DD HH:MM:SS-ZZZZ). For FIFA23, difficulty levels progress from easiest to most difficult as: *Beginner, Amateur, Semi-Pro, Professional, World Class, Legendary.* Street Fighter V difficulties are indicated numerically from *(1)* through *(8)*, where 1 is easiest and 8 is most difficult. The session number is reset for each unique participant-game-difficulty combination to track progression within specific difficulty levels.

"Splits": The Cross-Validation Folds

The structure consists of 7 outer folds, each is further divided into 7 inner folds. Each test set contains 4-6 participants, while validation sets comprise 3-5 participants. The splitting strategy ensures representation of minority classes across all folds to address data imbalance concerns. This nested structure supports various evaluation approaches: the outer folds provide unbiased performance estimates, while inner folds enable systematic hyperparameter tuning or ensemble model development. The consistent participant-level splitting across all folds ensures reproducible benchmarking for future research using this dataset.

"Human Panel Samples": The Human Evaluation Subset

As detailed in Table 3.6, the folder contains standardized webcam recordings and corresponding annotation data. The Samples folder contains 20 gameplay ses-

Table 3.5: Structure of Questionnaire Files

File	Column	Description
participants.csv	participant_id	Unique identifier
	age	Participant age
	sex	M/F
	fifa_exp	FIFA experience (0-4)
	sf_exp	Street Fighter experience (0-4)
submissions.csv	submission_id	Unique session identifier
	participant_id	Player identifier
	game	FIFA23 or Street Fighter V
	difficulty	Game-specific level
	session_no	Sequential session number
	start_ts	Session start
	end_ts	Session end timestamp
	engagement	Overall engagement (0-4)
	interest	Interest/enjoyment (0-4)
	stress	Stress level (0-4)
	excitement	Excitement level (0-4)

sions, with video files named using the format `<session_id>.mp4`. Each video maintains consistent quality specifications (480x480 pixels, 30 FPS) achieved by cropping the player webcam feed from the original OBS recordings. The Training folder contains five reference samples representing distinct engagement levels (two high, two neutral, one low), sourced from different participants to establish diverse baseline examples. These training samples were used to calibrate annotators and establish common rating criteria before their evaluation of the main sample set. The samples were selected to represent various engagement levels and player demographics, enabling comprehensive assessment of human annotators' rating consistency and accuracy. To facilitate evaluation against human annotators, the samples were sourced exclusively from the test set of the final outer fold (Fold 6).

Table 3.6: Structure of Human Panel Folder

Folder/File	Contents/Column	Description
Samples/	20 files	Webcam footage cropped
	session_id.mp4	from OBS recordings
Training/	high1, high2	Engagement labels 3-4
	neutral1, neutral2	Engagement label 2
	low1	Engagement labels 0-1
annotations.csv	participant_id	Unique player identifier
	submission_id	Unique session identifier
	engagement	Ground truth rating (0-4)
	annotator_[0-13]	Annotator ratings (0-4)
	annotator_[0-13]_conf	Annotator confidence (0-4)
annotators.csv	annotator_id	Unique identifier (0-13)
	experience	Gaming experience (0-4)
	clues	Engagement indicators used

3.3 Data Processing and Methodology

This section presents the methodology for the exploratory analysis of the MultiPENG dataset, investigating how engagement manifests across different sensing channels. As identified in our research directions, comprehensive multimodal datasets with reliable engagement annotations are crucial for understanding player engagement. We detail the data processing and model architecture we used for player engagement measurement.

3.3.1 Data Processing

To prepare the multimodal data for our engagement classification model, we implemented modality-specific processing pipelines that extract relevant features while preserving temporal structure. Our approach maintains the native sampling rates of each modality while addressing practical constraints through targeted downsampling. The downsampling rate we used was chosen through trial and error.

EEG. For EEG data, we used band powers (theta, alpha, low/high beta, gamma) from all 14 channels as provided by the EMOTIV software without additional filtering or artifact removal beyond the system's built-in processing. Our implementation filters out low-quality signals by setting band power values to null when the overall quality score falls below a configurable threshold, with interpolation applied to maintain sequence continuity. The data is then downsampled by taking every third sample to reduce sequence length.

Eye Tracking. For eye tracking, we extracted specific features from the Gazepoint software output including fixation position and duration (FPOGX, FPOGY, FPOGD), saccade characteristics (SACCADE_MAG, SACCADE_DIR), pupil diameter (LPD, RPD), and blink patterns (binary blink state, blink duration, and blinks per minute). After interpolating missing values, we downsample to every third sample, effectively reducing the 60Hz data to a more manageable rate.

Webcam. Facial features were extracted through two complementary methods. From OpenFace [144], following the approach in [40], we focused on head pose dynamics (3D translation and rotation vectors) and facial action unit intensities. We utilized the intensity values (ranging from 0 to 5) of all 17 action units detected by OpenFace and calculated derivative features including head velocity and acceleration by computing differences between consecutive frames. This data was downsampled to every third frame. Additionally, we extracted facial embeddings using EmoNet [104] by truncating the output layers.

Heart Rate. Heart rate data, while collected and included in the dataset, was not incorporated into our neural architecture. This decision was influenced by the known challenges in processing photoplethysmography (PPG) signals from consumer-grade wearables. The complexity of properly filtering and normalizing these signals, especially during physical movement associated with gameplay, presents an opportunity for future work with this dataset. Unfortunately, there

were many missing artifacts in our collected heart rate data, which prevented us from obtaining meaningful HR-based engagement classification results.
Across all modalities, we applied min-max scaling to normalize feature values, with scaling parameters determined exclusively from the training set to prevent data leakage. This standardized processing pipeline ensures reproducibility while maintaining the unique temporal characteristics of each modality. The detailed description of feature columns used in our methodology is shown in Appendix A.

3.3.2 Methodology

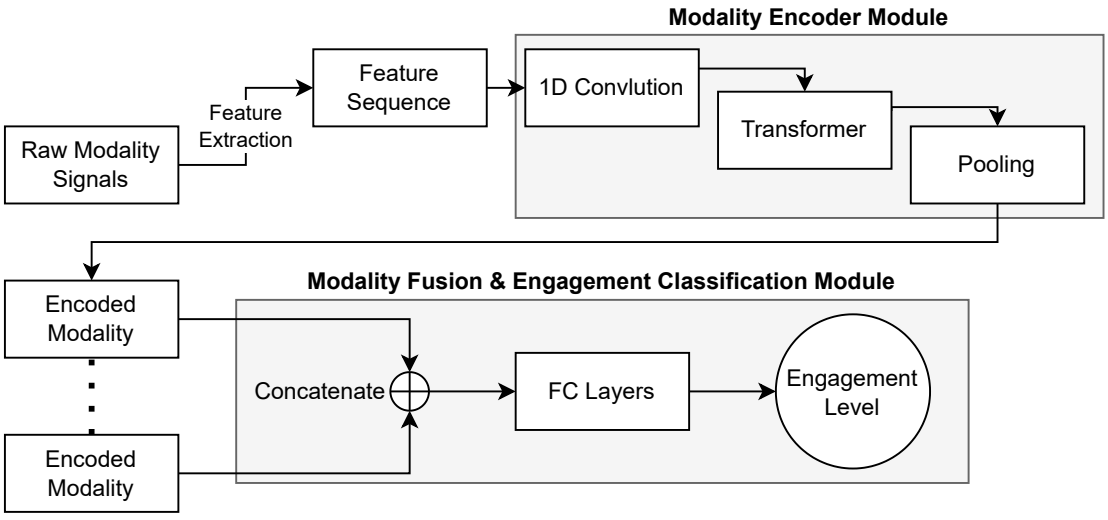
To effectively capture the temporal dynamics of engagement while handling the high dimensionality and varied sampling rates of our multimodal data, we developed a specialized neural architecture. The model balances complexity against our dataset size constraints to prevent overfitting while enabling effective feature extraction across modalities.

Model Architecture

At the core of our approach is a hybrid architecture that combines convolutional downsampling with transformer-based sequence modeling, as shown in Figure 3.5. This architecture is specifically designed to:

1. Efficiently process variable-length sequential data
2. Maintain modality-specific processing without information loss through resampling
3. Capture both local temporal patterns and long-range dependencies
4. Produce fixed-length representations suitable for fusion and classification

Figure 3.5: Multimodal engagement classification architecture. Each modality is processed by a separate ConvTransformerEncoder to extract modality-specific representations before late fusion.



The model consists of two main components:

ConvTransformerEncoder. This module processes each modality independently and consists of:

- **Convolutional Downsampling**: A 1D convolutional layer (kernel size 3, stride 2) that reduces sequence length while projecting the input features to a uniform hidden dimension. This serves as dimensionality reduction for high-dimensional inputs like the 60 Hz eye tracking signals while preserving temporal locality.
- **Transformer Encoder**: A lightweight transformer (1-2 layers, single attention head) that captures dependencies between timesteps after convolutional downsampling. This component is particularly important for modeling engagement's temporal dynamics, as engagement states often depend on patterns spanning multiple timesteps.
- **Global Pooling**: An adaptive average pooling layer that produces a fixed-length representation regardless of input sequence length, allowing the model to handle variable-length sequences without padding or truncation.

The encoder implements a form of hierarchical feature extraction, where local features are first captured by convolutions before modeling their relationships with transformers. This approach is computationally efficient while maintaining representational power.

MultimodalFusion. This module implements late fusion of modality-specific features:

- Independent modality processing through separate ConvTransformerEncoder instances
- Concatenation of the resulting feature vectors
- Projection to a unified hidden representation through a fully-connected layer
- Final binary classification through a sigmoid activation

Figure 3.6 shows how we implement the proposed framework using the two webcam-based modalities of OpenFace-calculated head pose and FAUs, and EmoNet-estimated facial embeddings. This example shows the justification of our architectural design choices in action. As a discovery phase in our research, this architecture is designed with the following key considerations:

- Prioritizing explainability: we need to be able to isolate modalities and explore their predictive power in engagement measurement and their contribution in multimodal integration.
- Fair cross-modal comparison: we need to maintain model consistency with different modalities for fair performance comparison.
- Adaptability: we need to handle modalities with different sampling rates and input shapes while avoiding one modality overwhelming others due to dimensionality or scale

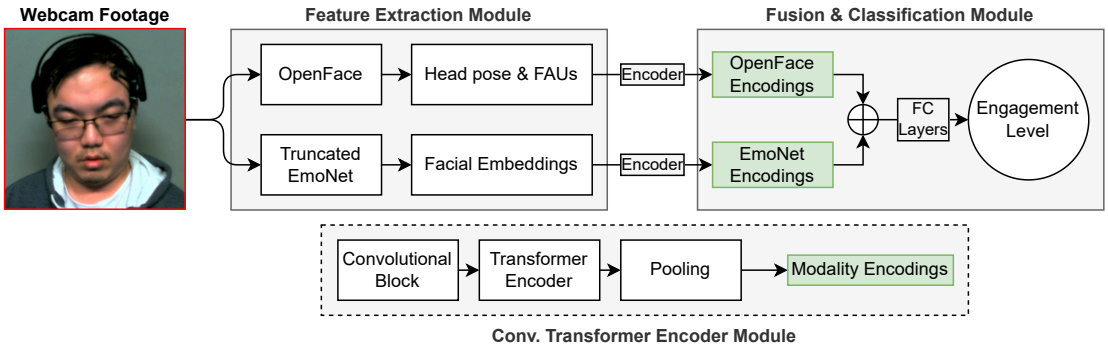


Figure 3.6: Proposed multimodal model implemented using two webcam-based modalities.

Beyond the discovery nature of this phase in our research, our framework provides higher robustness through late fusion. We chose late fusion over early fusion (concatenating raw features) or intermediate fusion (combining at intermediate layers) because it allows each modality to be processed optimally according to its characteristics before integration. This is particularly important given the diverse nature and granularities of our modalities (facial, eye tracking, EEG, and gameplay features). Additionally, having an encoder per modality preserves modality-specific patterns, allowing for a graceful handling of missing modality data, asynchronous learning across modalities (i.e. one modality might converge faster than another), and efficient inference through potentially skipping some modalities when confident. Furthermore, this framework allows explicitly controlling which modality should contribute to the final prediction. This allows for learning attention weights or gating mechanisms to dynamically balance modalities based on the input. However, we processed modalities uniformly to fairly analyze their performance contribution.

Training Configuration

The model was trained with the following configuration:

- **Optimization**: Adam optimizer with learning rate 5e-5
- **Loss Function**: Binary Cross-Entropy with class weighting to address imbalance
- **Class Weighting**: $\text{pos\_weight} = \frac{\text{count(negative\_samples)}}{\text{count(positive\_samples)}}$
- **Batch Size**: 64 samples per batch
- **Training Regime**: 150 epochs maximum with early stopping (patience: 10 epochs)
- **Cross-Validation**: Ensuring complete separation between participants in training, validation, and test sets

Our training approach was designed to address several challenges specific to our dataset:

Limited Sample Size. With approximately 580 training samples distributed across multiple participants and varied engagement levels, overfitting was a significant concern. We employed multiple strategies to mitigate this:

- Regularization through dropout (0.3-0.5)
- Early stopping based on validation performance
- Relatively small model size (hidden dimension of 128, single attention head)
- Convolutional downsampling to reduce the effective sequence length

Handling Class Imbalance. To address measurement uncertainty in self-reported engagement and the practical limitations of our dataset, we transformed the Likert responses into binary classifications using the mean reported engagement (2.87) as the threshold. While regression approaches preserve the full granularity of engagement scores and can capture subtle improvements over time, they require larger sample sizes to reliably model the continuous relationship and are particularly sensitive to the systematic response biases evident in our highly skewed data, as shown in Table 3.3. The binary transformation provides a more robust classification framework that accounts for person-specific reporting tendencies through group-stratified cross-validation, while offering meaningful discrimination between participants with above- and below-average engagement patterns despite the reduced granularity. To further address the resulting class imbalance (265 low vs. 635 high engagement samples), we applied inverse frequency weighting in the loss function to prevent the model from being biased toward the majority class.
We deliberately avoided synthetic data generation techniques such as SMOTE or generative augmentation approaches. While these methods can address class imbalance, they risk introducing modality-specific artifacts that could unfairly advantage certain input types over others. Since our primary objective is comparing the discriminative power of different modalities, maintaining the authentic statistical properties of each modality is essential for fair cross-modal evaluation. The class weighting approach preserves the original data distribution while ensuring balanced learning across engagement levels.

Modality-Specific Processing. Different modalities in our dataset have varying sampling rates: webcam footage at 30Hz, EEG band data at 8Hz, and eye tracking at 64Hz. Rather than resampling to a common rate (which would lose information in higher-frequency signals or introduce redundancy in lower-frequency ones), we processed each modality at its native sampling rate. This approach preserves the fine-grained temporal information present in each modality. We empirically validated this decision by comparing against models trained on resampled data (at 10Hz and 20Hz), finding that native-rate processing consistently outperformed resampled approaches.

Cross-Validation Strategy. We employed a participant-based cross-validation scheme with consistent inner splits. For each outer fold (test set), we used a single fixed train/validation split rather than multiple inner fold combinations. The validation set served dual purposes: hyperparameter optimization using Optuna

and early stopping during training. Hyperparameters optimized included learning rate, weight decay, batch size, model architecture parameters (number of attention heads, projection dimensions, hidden layer size, number of layers), and dropout rate. This approach ensures each test participant is evaluated exactly once with a model trained on a consistent training subset. By providing these predefined splits in the dataset, we establish a standardized evaluation framework that enables fair comparison of different modeling approaches while maintaining the integrity of participant-based separation between train, validation, and test sets. Using a single inner fold per test set is sufficient for our analysis, as exhaustive inner fold evaluation would add unnecessary computational overhead without providing additional insights about model generalization.

Sequence Processing. The combination of convolutional and transformer-based processing allows the model to handle the long sequences (hundreds of timesteps) common in our data without requiring fixed-length inputs and with less susciptibility to overfitting. This flexibility is particularly important for comparing engagement levels across gameplay sessions of different durations.
This architecture represents a balance between complexity and practicality for engagement classification with multimodal time series data. The design allows direct comparison between different sensing modalities by using identical architectures with modality-specific parameters, facilitating fair evaluation of each modality's contribution to engagement estimation.

3.4 Results and Discussion

In this section, we present the results for our engagement estimation model. First, we systematically evaluate individual modalities—facial expressions, eye metrics, and EEG signals—to determine their relative effectiveness for engagement prediction, followed by an analysis of multimodal integration and a flow theory-based approach. We then compare our computational approaches with human annotation performance to establish contextual benchmarks for engagement detection. Our discussion is enriched with a comprehensive analysis of the models used, which will inform our subsequent development of non-intrusive, real-time measurement techniques in complex gaming environments.
Throughout this section, all reported uncertainties ($\pm$) represent the standard error of the mean (SEM) of the corresponding metric. The SEM quantifies the precision of the sample mean and is calculated by dividing the standard deviation (σ) by the square root of the number of observations (n), which in our case are the metric scores across different cross-validation folds:

$$\text{SEM} = \frac{\sigma}{\sqrt{n}} \tag{3.1}$$

3.4.1 Player Engagement Estimation Results

To evaluate the effectiveness of different engagement measurement approaches, we compared performance across individual sensing modalities and multimodal fusion techniques. Table 3.7 presents the detailed classification metrics for each

approach.

Table 3.7: Classification Performance Comparison Across Different Approaches

Metric	Low	High	Avg.	ROC_AUC	Accuracy
Webcam (OpenFace and EmoNet)					
Precision	0.32 ± 0.07	0.76 ± 0.03	0.54 ± 0.04		
Recall	0.43 ± 0.10	0.69 ± 0.08	0.56 ± 0.03	0.56 ± 0.04	0.63 ± 0.04
F1-score	0.35 ± 0.08	0.71 ± 0.04	0.53 ± 0.03		
Eye					
Precision	0.18 ± 0.07	0.75 ± 0.05	0.46 ± 0.05		
Recall	0.33 ± 0.15	0.64 ± 0.13	0.48 ± 0.02	0.48 ± 0.04	0.59 ± 0.06
F1-score	0.22 ± 0.09	0.63 ± 0.09	0.43 ± 0.02		
EEG					
Precision	0.35 ± 0.09	0.78 ± 0.05	0.57 ± 0.05		
Recall	0.42 ± 0.15	0.64 ± 0.12	0.53 ± 0.03	0.57 ± 0.06	0.61 ± 0.05
F1-score	0.31 ± 0.09	0.65 ± 0.07	0.48 ± 0.02		
Multimodal (EEG + EYE + OpenFace + EmoNet)					
Precision	0.40 ± 0.07	0.76 ± 0.02	0.58 ± 0.03		
Recall	0.41 ± 0.09	0.73 ± 0.06	0.57 ± 0.03	0.59 ± 0.04	0.65 ± 0.03
F1-score	0.38 ± 0.07	0.73 ± 0.04	0.56 ± 0.03		
Flow-Based Model (Skill and Challenge)					
Precision	0.42 ± 0.05	0.78 ± 0.03	0.60 ± 0.03		
Recall	0.49 ± 0.08	0.74 ± 0.03	0.62 ± 0.03	0.63 ± 0.03	0.67 ± 0.02
F1-score	0.44 ± 0.06	0.76 ± 0.02	0.60 ± 0.03		

Signal Quality Threshold Analysis

Before comparing different engagement estimation approaches, we conducted a systematic analysis to determine optimal signal quality thresholds, particularly for EEG data which is highly susceptible to noise and motion artifacts during gameplay. Figure 3.7 illustrates the relationship between minimum EEG quality thresholds and classification performance.

As shown in the figure, classification performance initially improves as the quality threshold increases, reaching optimal performance at approximately 50%. This pattern suggests that excluding low-quality signals up to this point effectively removes noise that would otherwise confound the engagement classification. The sharp performance increase between thresholds of 0% and 50% indicates that a substantial portion of the EEG signal contains noise that negatively impacts classification accuracy when included, confirming the findings of Figure 3.3.

Interestingly, performance plateaus and then slightly decreases at thresholds above 50%, suggesting that overly strict quality requirements may eliminate informative but imperfect signals. The relatively stable performance between thresholds of 40% and 100% (with fluctuations within the margin of error) indicates that the classifier can effectively handle some interpolated data without significant performance degradation.

Based on this analysis, we established 50% as the optimal EQ.OVERALL threshold for our EEG data processing pipeline, balancing noise reduction against data

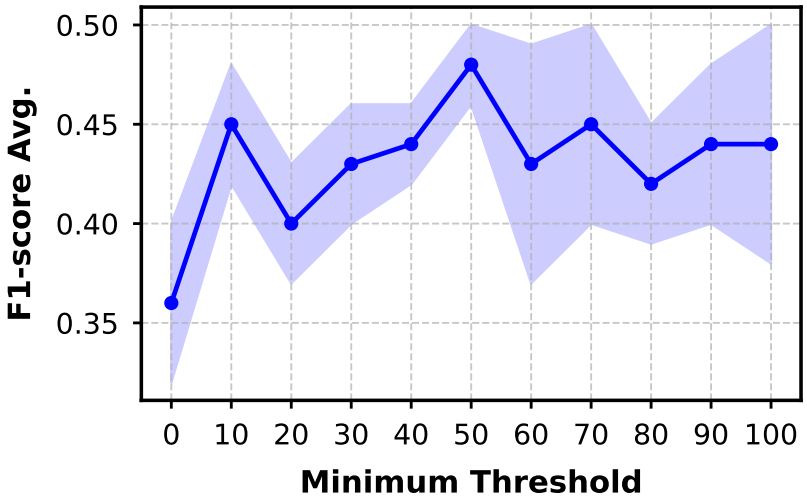


Figure 3.7: Impact of EEG quality threshold on classification performance. The x-axis shows the minimum EQ.OVERALL threshold, while the y-axis shows the resulting F1-score average across classes. The shaded area represents the standard error (SEM).

availability. This threshold was used for all EEG-based classification results reported in Table 3.7.

While our analysis identified 50% as the optimal static threshold for the current dataset, it is important to acknowledge that this threshold was derived from a relatively small participant pool with approximately 20 data points per individual, making the results potentially sensitive to participant-specific characteristics. In practice, an adaptive or dynamic thresholding approach that adjusts quality requirements based on individual participant signal characteristics or real-time signal conditions could provide more robust performance across diverse populations and recording environments. However, implementing such adaptive mechanisms would introduce additional complexity in both algorithm design and validation, requiring more extensive datasets to properly characterize individual differences and establish reliable adaptation rules. Given our current study's scope and the practical constraints of real-time gameplay applications where computational simplicity is advantageous, we opted to use the empirically determined static threshold of 50%, while recognizing that future work with larger, more diverse participant samples would benefit from investigating adaptive quality assessment approaches.

Unimodal Performance Analysis

Among individual sensing modalities, the webcam-based approach (combining OpenFace structural analysis and EmoNet embeddings) achieved the highest accuracy at 63% (±0.04). This modality demonstrated particularly strong performance in identifying high engagement states (precision: 0.76 ± 0.03, recall: 0.69 ± 0.08), suggesting that facial expressions and head movements provide valuable indicators of engagement. The webcam approach offers significant practical advantages as it requires no specialized hardware beyond standard computing equipment, making it highly accessible for real-world applications.

EEG features derived from band powers across different frequency ranges yielded the second-highest unimodal accuracy at 61% (±0.05) when using the established 50% quality threshold. While EEG showed the highest precision for detecting high engagement states (0.78 ± 0.05) among all unimodal approaches, its relatively lower recall (0.64 ± 0.12) indicates that it may miss some instances of high engagement. The quality threshold analysis was particularly important for EEG processing, as performance varied considerably with different threshold settings (as shown in Figure 3.7). Despite its strong performance after quality filtering, the specialized equipment, calibration requirements, and need for signal quality management present significant barriers to practical implementation in everyday gaming scenarios.

Eye tracking features, including fixations, saccades, pupil diameter, and blink patterns, achieved the lowest accuracy among tested modalities at 59% (±0.06). The computational analysis of eye movements appears to capture engagement patterns that may not be immediately obvious through visual observation alone.

Multimodal Integration Benefits

Our multimodal approach combining EEG, eye tracking, and facial features achieved 65% accuracy (±0.03), demonstrating the value of integrating complementary sensing channels. The fusion model showed balanced performance improvements across metrics compared to individual modalities, with notable gains in low engagement detection (precision: 0.40 ± 0.07 compared to 0.32 ± 0.07 for webcam alone). This improvement suggests that different modalities may capture complementary aspects of the engagement experience, with their combination providing a more comprehensive measurement framework.

However, the practical significance of this improvement must be considered against the substantial increase in implementation complexity. The multimodal approach requires synchronization of multiple specialized sensors and significantly more computational resources for feature extraction and processing. The marginal 2 percentage point improvement over the webcam-only approach may not justify these additional requirements in many real-world applications.

Flow Theory-Based Model

Building on the Flow Theory, which suggests optimal engagement occurs when skill matches challenge levels, we trained an SVM classifier using participant experience as a skill proxy and normalized game difficulty as a challenge measure, with balanced class weights. The model incorporated two key features: player skill (represented by self-reported experience levels on a 5-point scale) and normalized game difficulty (6 levels for FIFA'23, 8 levels for SFV, normalized to 0-1 range).

Perhaps the most interesting finding is that our flow theory-based model, using only player skill and game challenge as predictors, achieved the highest overall accuracy at 67% (±0.02). This relatively simple approach slightly outperformed even our complex multimodal neural architecture while requiring substantially less computational overhead and no specialized sensing equipment.

As shown in Table 3.7, the flow-based model demonstrated particularly strong performance in detecting high engagement states (precision: 0.78 ± 0.03, recall: 0.74 ± 0.03) and offered the most balanced performance for low engagement de-

tection (precision: 0.42 ± 0.05, recall: 0.49 ± 0.08) among all approaches tested. This balanced performance is reflected in its superior F1-scores for both low (0.44 ± 0.06) and high (0.76 ± 0.02) engagement classes.
Examining the ROC-AUC scores reveals varying levels of discriminative performance across approaches. The flow-based model achieves the highest ROC-AUC at 0.63 (±0.03), followed by the multimodal (0.59 ± 0.04), EEG (0.57 ± 0.06), and webcam (0.56 ± 0.04) approaches, with the eye-based approach showing the lowest performance at 0.48 (±0.04). The flow-based model's ROC-AUC of 0.63, while the best among the tested approaches, indicates moderate discriminative ability that falls short of strong classification performance (typically >0.7). However, this score aligns reasonably well with the model's accuracy of 67%, suggesting that the model produces adequately calibrated probabilities for threshold-based decision making, though there remains room for improvement in separating the two classes across the full probability range.
The multimodal approach provides the highest ROC-AUC, indicating better discrimination across different thresholds. This characteristic makes it potentially more valuable in applications requiring fine-grained control over the sensitivity/specificity balance, even if its maximum accuracy is slightly lower than the flow-based model.
These results align with Csikszentmihalyi's flow theory[7], which suggests that optimal engagement occurs when skill and challenge are balanced. The model's strong performance with just these two features provides compelling empirical support for the theory's practical application in engagement estimation. However, it's important to note that this approach depends on having explicit measures of player skill and game difficulty, which may not be readily available in all gaming contexts.

3.4.2 Human Annotation Analysis

To contextualize our computational approaches, we conducted a human annotation analysis to understand the capabilities and limitations of humans in visually assessing engagement in gaming contexts. We recruited 14 human judges who, after training on 5 reference samples, analyzed webcam footage from 20 gameplay sessions across 5 different participants. Judges were provided with sample videos representing different engagement levels and suggestions for potential visual cues (e.g., eye blinking patterns, head movements, facial expressions) without mandating specific indicators.
Inter-rater agreement analysis revealed consistently low agreement across different granularities: raw 5-point Likert scores (Krippendorff's $\alpha = 0.092$, Cohen's $\kappa = 0.001$), Low/Neutral/High classes ($\alpha = 0.080$, $\kappa = 0.036$), and binary Low/High classes (raw agreement: 52.1%, $\alpha = 0.040$, $\kappa = 0.043$). When evaluated against survey answers, human annotators achieved 50±03% accuracy, with precision, recall, and F1-scores all around 0.5, as shown in Table 3.8. This uniformly low agreement across metrics highlights a fundamental challenge: unlike explicitly manifested emotional states (e.g., happiness indicated by smiling), engagement appears to lack consistently interpretable visual cues. Notably, even participants could not reliably identify their own previous engagement states when reviewing their recordings.

Table 3.8: Performance Comparison on Human Annotation Subset (20 Samples)

Metric	Low	High	Avg.	ROC_AUC	Accuracy
Human Annotators					
Precision	0.59 ± 0.03	0.38 ± 0.04	0.49 ± 0.03		
Recall	0.53 ± 0.05	0.45 ± 0.06	0.49 ± 0.03	-	0.50 ± 0.03
F1-score	0.55 ± 0.04	0.40 ± 0.04	0.47 ± 0.03		
Webcam - No Ensemble					
Precision	0.54 ± 0.12	0.42 ± 0.02	0.48 ± 0.07		
Recall	0.26 ± 0.09	0.77 ± 0.06	0.51 ± 0.03	0.50 ± 0.03	0.46 ± 0.04
F1-score	0.32 ± 0.09	0.53 ± 0.02	0.43 ± 0.05		
Webcam - Ensemble					
Precision	0.50	0.38	0.44		
Recall	0.17	0.75	0.46	0.46	0.40
F1-score	0.25	0.50	0.38		
Multimodal - No Ensemble					
Precision	0.73 ± 0.13	0.41 ± 0.03	0.57 ± 0.07		
Recall	0.38 ± 0.12	0.79 ± 0.09	0.58 ± 0.03	0.61 ± 0.04	0.52 ± 0.05
F1-score	0.44 ± 0.12	0.52 ± 0.02	0.48 ± 0.06		
Multimodal - Ensemble					
Precision	0.83	0.42	0.63		
Recall	0.42	0.83	0.63	0.63	0.56
F1-score	0.56	0.56	0.56		

To compare human and computational performance, we evaluated our machine learning models on the same 20 samples that humans annotated. These samples came from 5 participants whose data was completely excluded from the training and validation sets, maintaining strict separation between training and testing data. We conducted two evaluations: applying each of the 7 fold-specific models independently and reporting the average performance, and implementing an ensemble approach where predictions from all 7 models were averaged. We selected two approaches for comparison: the webcam-based model (processing the same visual information available to human judges) and the multimodal approach (incorporating all sensing modalities).

As shown in Table 3.8, several notable patterns emerge. The multimodal ensemble approach achieves the highest performance (56% accuracy, 0.63 ROC-AUC), modestly outperforming human annotators (50% accuracy). Comparing ensemble to non-ensemble results reveals the benefit of the ensemble approach, with the multimodal ensemble showing a 4 percentage point improvement over the average individual model (56% vs 52%). This improvement demonstrates that aggregating predictions across multiple models helps mitigate individual model biases and increases robustness, particularly on challenging samples. However, this performance still falls below the same model's capabilities on the full dataset (65% accuracy), suggesting these particular samples present specific classification challenges. This substantial drop in performance (from 65% to 56%) indicates significant variability in classification difficulty among samples, with this subset representing particularly challenging cases.

The webcam-based model, despite using the same visual information available to human judges, underperforms compared to human annotators (40-46% accuracy vs. 50%). This suggests that human observers may leverage contextual cues or implicit knowledge not captured by computational analysis of facial expressions and head movements alone. The performance gap between the webcam-based approach and human judges, contrasted with the multimodal approach's superior performance, suggests that the webcam model might require a larger training dataset to properly generalize.

The multimodal approach demonstrates improved discrimination ability, particularly for low engagement states (precision of 0.83 for the ensemble model compared to 0.59 for human annotators). The fact that augmenting webcam-based features with eye tracking and EEG data enables the model to outperform human judges indicates room for improvement in the webcam-only approach. Since eye tracking features can potentially be estimated from webcam data, this suggests that with sufficient training data and fine-tuning, a webcam-based approach might eventually match or exceed human performance.

Examining the ROC-AUC scores provides additional insights into model discrimination capabilities. The multimodal approach achieves the highest ROC-AUC (0.63 for ensemble), indicating better overall class separation ability than the webcam-based approach (0.46 for ensemble). However, both computational approaches show class imbalance in their predictions—the webcam model favors high engagement detection (recall of 0.75-0.77 for high vs. 0.17-0.26 for low engagement), while the multimodal approach, particularly in its ensemble form, achieves a more balanced detection capability (precision of 0.83 for low and recall of 0.83 for high engagement).

These findings establish several important conclusions. First, engagement measurement remains challenging even for computational approaches when evaluating specific participant subsets. Second, engagement manifests across multiple physiological and behavioral channels rather than through any single observable modality, explaining why the multimodal approach outperforms the webcam-only method. Finally, the consistently low inter-rater agreement among human judges confirms engagement's nature as a complex internal experience rather than an easily observable state.

The key question emerging from this analysis is why the webcam-based model underperforms relative to human judges on these specific 20 samples, while showing stronger performance on the full dataset. Notably, the multimodal approach maintains more consistent performance across both the full dataset and the subset, whereas the webcam-based model shows a marked drop on these specific samples. This pattern suggests there could be participant-specific sensitivity that the webcam-based approach is more susceptible to than other methods. It also raises important questions about the statistical power of our dataset, which we address in the following section through a systematic power analysis of both the multimodal and webcam-based approaches.

3.4.3 Statistical Power Analysis

Before analyzing the effectiveness of different modalities for engagement prediction, we conducted a statistical power analysis to ensure our dataset contained

sufficient samples to detect meaningful differences in model performance. This analysis is particularly important given the relative complexity of our models and the challenges inherent in measuring subjective psychological states like engagement.

We focused on two key performance metrics: F1-score (average) and ROC-AUC. These metrics were selected over accuracy due to their robustness to class imbalance and ability to capture both precision and recall trade-offs, which is particularly important given the imbalanced nature of our engagement classes.

Our power analysis employed a methodical approach using Cohen's d effect size to quantify the magnitude of performance differences between models trained on subsets of varying sample sizes (n=5, 10, 15, 20) compared to models trained on the full training dataset. This approach follows established methods for evaluating sample size effects in machine learning applications [146].

Cohen's d represents the standardized difference between two means, calculated as:

$$d = \frac{|\mu_1 - \mu_2|}{\sqrt{\frac{\sigma_1^2 + \sigma_2^2}{2}}} \tag{3.2}$$

Where μ_1 and μ_2 are the means of the performance metrics, and σ_1 and σ_2 are the standard deviations.

We calculated statistical power using the non-central t-distribution approach:

$$\text{Power} = 1 - P(t < t_{\text{crit}} | \text{df}, \text{ncp}) \tag{3.3}$$

Where t_{crit} is the critical t-value at $\alpha = 0.05$, df is the degrees of freedom (2n - 2), and ncp is the non-centrality parameter, calculated as $d \cdot \sqrt{\frac{n}{2}}$.

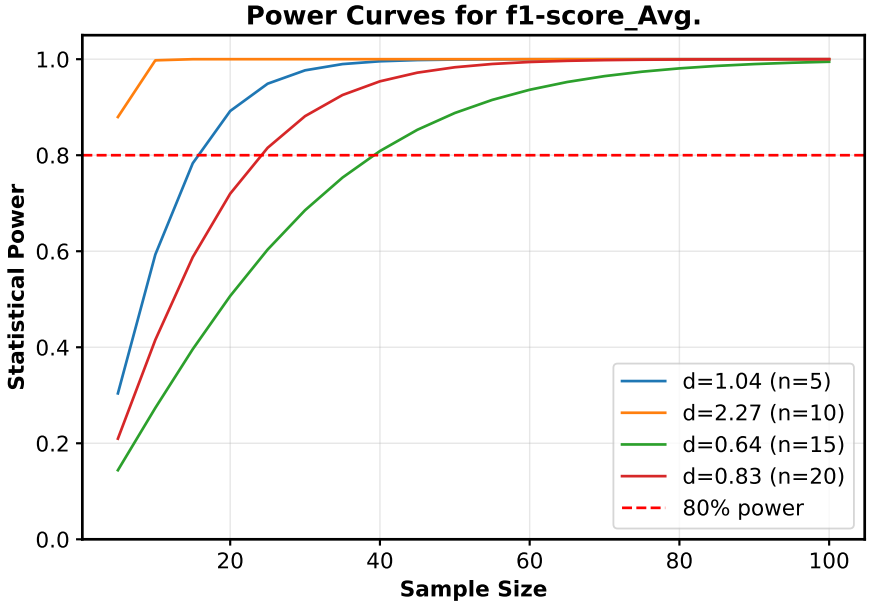


Figure 3.8: Power curves for F1-score average. The curves show statistical power as a function of sample size for different effect sizes observed in our cross-validation multimodal experiments. The horizontal dashed line represents the conventional 80% power threshold.

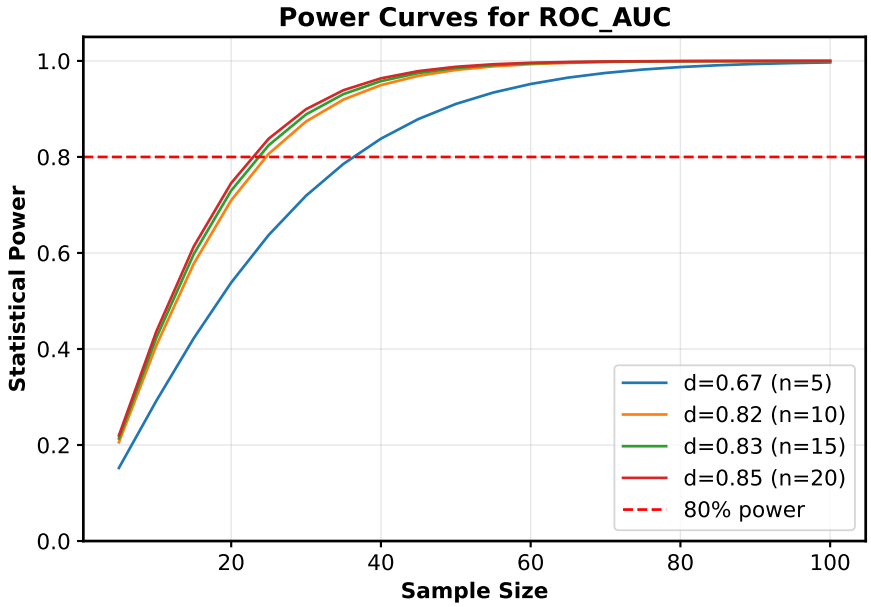


Figure 3.9: Power curves for ROC-AUC. The curves show statistical power as a function of sample size for different effect sizes observed in our cross-validation multimodal experiments. The horizontal dashed line represents the conventional 80% power threshold.

A notable observation in our analysis was the substantial fluctuation in effect sizes (Cohen's d) across different sample sizes. For F1-score, effect sizes varied considerably: n=5: d=1.04 (large effect); n=10: d=2.27 (very large effect); n=15: d=0.64 (medium effect); and n=20: d=0.83 (large effect). ROC-AUC showed more stability but still varied: n=5: d=0.67; n=10: d=0.82; n=15: d=0.83; and n=20: d=0.85.

These fluctuations, particularly in the F1-score metric, highlight what recent literature refers to as the "cold-start problem" in machine learning evaluation[147]—the challenge of model performance instability when adding new participants, especially in the early stages of data collection. The dramatic increase in F1-score effect size from n=5 to n=10, followed by a substantial drop at n=15, suggests that model performance is highly sensitive to the specific participants included in the analysis. This non-linear pattern reflects the complexity of modeling engagement across different individuals with varying gameplay styles and emotional responses.

To validate that these observations were not specific to the multimodal approach, we conducted a parallel power analysis for the webcam-based model. Interestingly, this unimodal approach exhibited similar fluctuation patterns, with effect sizes of: n=5: d=0.96 (large effect, sufficient power); n=10: d=0.75 (medium-large effect, almost sufficient power); n=15: d=0.89 (large effect, sufficient power); n=20: d=0.76 (medium-large effect, almost sufficient power); and for the full dataset: d=0.65 (medium effect, insufficient power). These results confirm that the sensitivity to participant composition is not limited to complex multimodal approaches but extends to simpler unimodal models as well. The persistent fluctuations across both model types suggest that engagement measurement fundamentally involves complex individual differences that challenge consistent modeling across different

participant subsets, regardless of the sensing approach employed.
The analysis revealed differential power profiles across metrics, as illustrated in Figures 3.8 and 3.9. For F1-score average, the power eventually reached 0.9976, exceeding the conventional threshold of 0.8. While this high power value suggests statistical robustness, we must interpret it cautiously given the observed effect size fluctuations. The high power indicates that with our current sample size, we can detect differences in F1-score, but the magnitude and direction of these differences may be highly participant-dependent. In contrast, the power for ROC-AUC reached 0.7456, falling slightly below the conventional threshold of 0.8. This indicates near-sufficient but not optimal statistical power for detecting meaningful differences in ROC-AUC. Despite the more stable effect sizes observed for ROC-AUC across different sample sizes, the final power achieved was lower than expected. This suggests that ROC-AUC measurements may have higher variance in our dataset, potentially due to the complexity of engagement classification across diverse participants.
The difference in power between F1-score and ROC-AUC metrics is noteworthy and can be attributed to metric variability (ROC-AUC typically exhibits higher variance across cross-validation folds than F1-score, especially with imbalanced data) and effect size differences. Figure 3.9 shows that the power curves for ROC-AUC rise more gradually, requiring larger sample sizes to achieve the same power level compared to F1-score, particularly for smaller effect sizes.
These results suggest that while our sample size appears statistically robust for primary analyses based on F1-scores, the substantial fluctuations in effect sizes across different sample compositions indicate that our models' performance is sensitive to the specific participants included. This sensitivity highlights the inherent challenges in modeling subjective psychological states like engagement across different individuals. Findings related to ROC-AUC should be interpreted with additional caution due to the sub-optimal power achieved.
The power analysis ultimately demonstrates both the strengths and limitations of our dataset. It contains sufficient samples to support general conclusions about modality effectiveness, particularly regarding the comparison of different sensing approaches and the effectiveness of Flow Theory-based metrics for engagement prediction. However, the observed fluctuations in effect sizes emphasize the importance of considering individual differences when interpreting and applying these findings to new contexts or participants.

3.4.4 Ablation Study

To better understand the relative importance of each sensing channel within our multimodal approach, we conducted an ablation study that systematically evaluated performance impact when removing each modality. This analysis both quantifies unique contributions and provides further support for prioritizing webcam-based approaches.
Our experimental design established a baseline using the full model incorporating all four modalities (EEG, EYE, OpenFace, and EmoNet) trained across 7 cross-validation folds. We then created four variant models, each with one modality removed, and measured the resulting performance drop in terms of F1 score and ROC AUC.

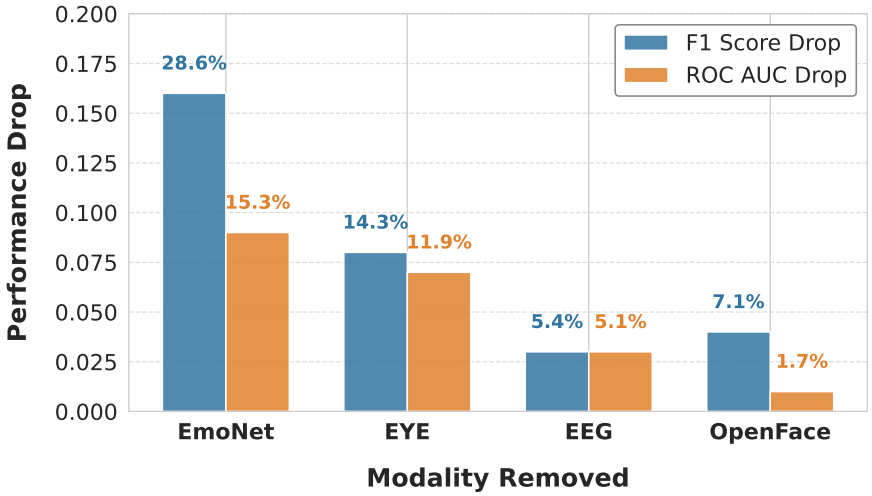


Figure 3.10: Performance drop when removing individual modalities from the multimodal model. The percentage values indicate the relative performance reduction compared to the full model.

As shown in Figure 3.10, the ablation results reveal a clear hierarchy in modality importance. Webcam-derived features, particularly EmoNet, provide the most substantial contribution, with its removal causing a 28.6% reduction in F1 score and a 15.3% drop in ROC AUC. This indicates that EmoNet's facial embeddings capture crucial engagement information not present in other modalities. Eye tracking emerged as the second most influential modality, with its removal resulting in a 14.3% decrease in F1 score and an 11.9% reduction in ROC AUC. OpenFace features and EEG signals showed more modest contributions, with their removal causing 7.1% and 5.4% drops in F1 score, respectively. The smaller impact of removing these modalities suggests redundancy with other channels—particularly notable for EEG, the most technically complex modality to implement. This indicates much of its engagement-relevant information may be accessible through other, more practical sensing channels.

Importantly, although eye tracking emerged as the second most important modality, many eye features (blink detection, gaze direction, pupil metrics) could potentially be estimated using webcam-based computer vision approaches. We deliberately avoided including such derived features in our OpenFace implementation to minimize inter-feature correlation, but this suggests a sophisticated webcam-based approach might capture substantial information from both facial and eye tracking modalities using a single sensor. This alignment between modality contribution and sensor practicality reinforces our earlier finding that webcam-based approaches achieved the highest unimodal accuracy.

These results have significant implications for engagement measurement system design. The dominant contribution of webcam-derived features strongly supports webcam-based approaches as offering an excellent balance of accuracy and accessibility. While these modalities contain overlapping information that cannot be simply summed, the potential for extracting eye-related features from webcam data strengthens the case for prioritizing webcam processing—particularly through advanced embedding techniques like EmoNet.

For applications with limited computational or data collection resources, focus-

ing on webcam processing would preserve much of the model's performance while significantly reducing implementation complexity. Future feature engineering efforts might be most productively directed toward improving facial embedding techniques and expanding webcam-based eye tracking capabilities, leveraging a single, widely available sensor to capture multiple engagement signals.

3.4.5 Feature Attribution Analysis

To gain deeper insights into the multimodal engagement model's decision-making process, we conducted a comprehensive feature attribution analysis using Captum [148], a model interpretability library for PyTorch. This analysis quantifies the contribution of individual input features to the model's predictions, helping to identify which modalities and specific signals most strongly influence engagement classification.

Methodology

We applied integrated gradients [149], a gradient-based attribution method that assigns importance scores to input features by integrating the gradients of the model's output with respect to its inputs along a straight-line path from a baseline (i.e. zero values) to the input. For each feature, we computed:

- Positive attributions (`pos_attr`): Contribution of the feature toward predicting high engagement
- Negative attributions (`neg_attr`): Contribution of the feature toward predicting low engagement
- Overall importance: Sum of the absolute values of positive and negative attributions ($|\texttt{pos\_attr}| + |\texttt{neg\_attr}|$)

This approach provides a nuanced view of feature importance, distinguishing between features that serve as indicators of high engagement, low engagement, or both. The analysis was performed across all four modalities (OpenFace, EmoNet, eye tracking, and EEG) to facilitate direct comparison of their relative contributions to engagement classification.

Results

Figure 3.11 presents the top 10 features ranked by overall importance. Notably, facial and eye tracking features dominate this ranking, with no EEG features appearing among the most important signals. The head pose translation along the z-axis (`pose_Tz`)—representing forward-backward head movement—emerged as the single most influential feature, followed by its acceleration (`acceleration_pose_Tz`). Pupil dilation metrics for both eyes and blink rate also ranked highly, while five of the top features came from EmoNet facial embeddings.
Examining features specifically associated with high engagement (Figure 3.12), we observe that facial embeddings from EmoNet constitute 7 of the top 10 features. The z-axis head pose translation remains prominent, suggesting that certain head positions strongly indicate engagement. Interestingly, theta band activity from the

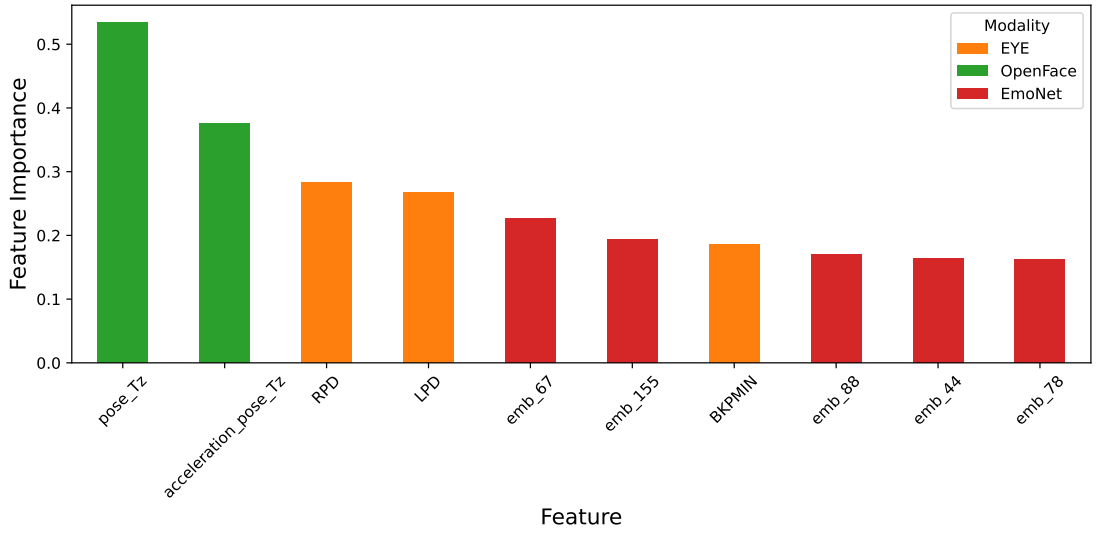


Figure 3.11: Top 10 features ranked by overall importance (absolute attribution)

AF3 electrode (`AF3.Theta`) appears among the top features positively associated with engagement, representing the only EEG feature in this ranking. This aligns with previous research by [32] that identified the ratio of theta activity at AF3 to alpha activity at P7 as an engagement index during gameplay. Right pupil dilation also shows strong positive attribution, consistent with literature linking pupil responses to cognitive engagement[36].

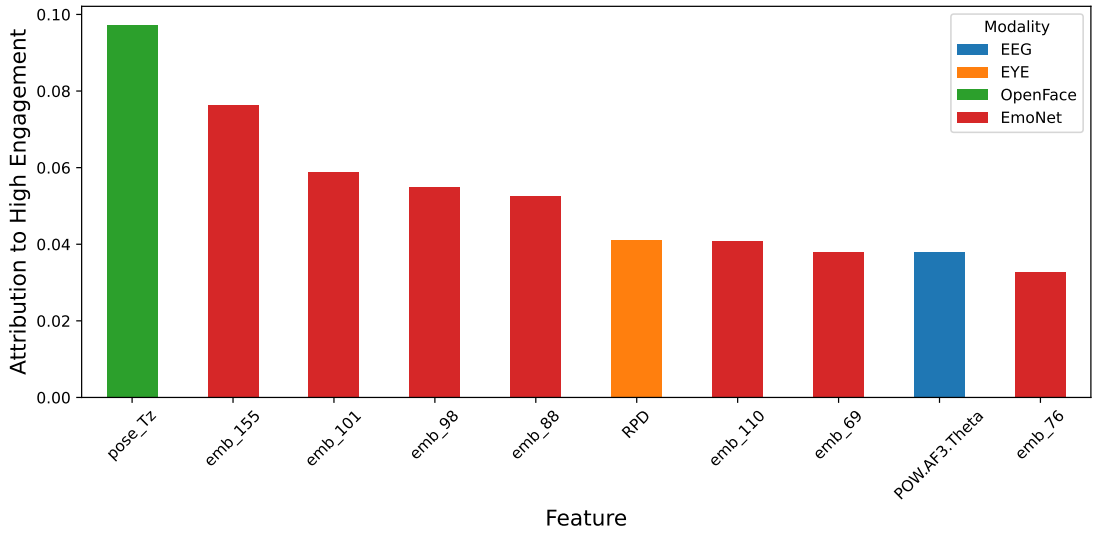


Figure 3.12: Top 10 features with highest positive attribution (indicating high engagement)

For features associated with low engagement (Figure 3.13), we again find `pose_Tz` and `acceleration_pose_Tz` as leading indicators. This suggests that head position and movement along the z-axis have complex relationships with engagement, with certain patterns indicating engagement and others indicating disengagement. Eye metrics including pupil dilation and blink rate appear strongly associated with low engagement predictions, as do several EmoNet facial embeddings. The F8 theta band activity represents the only EEG feature in this ranking. This is

particularly noteworthy as frontal theta activity has been established as an index of cognitive engagement and effortful processing [96], and is a key component in engagement indices such as Frontal Theta/Parietal Alpha that are sensitive to cognitive workload during gaming tasks [97]. Its appearance in the negative attribution ranking suggests that specific patterns of right frontal theta activity may indicate disengagement in certain gaming contexts.

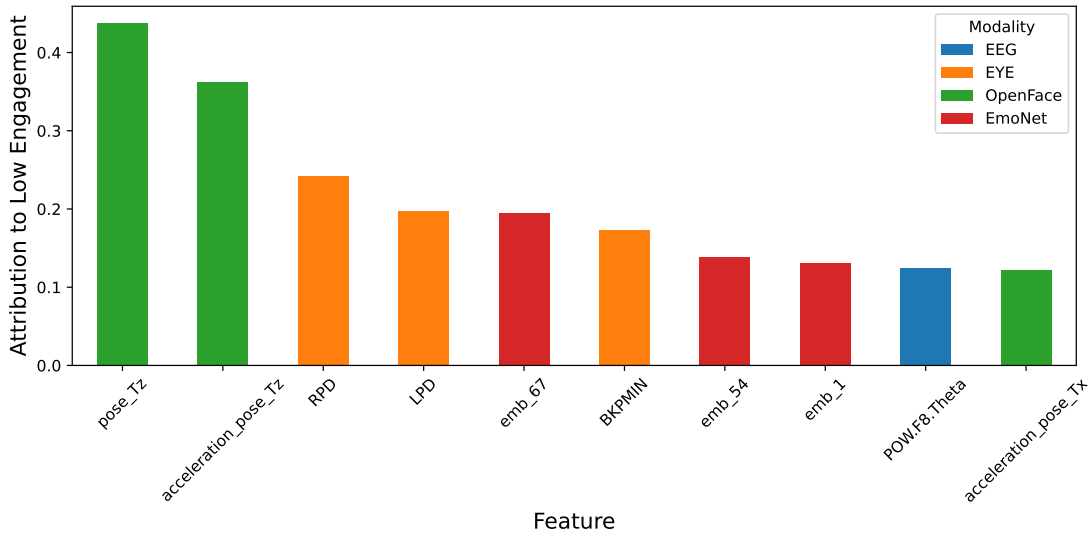


Figure 3.13: Top 10 features with highest negative attribution (indicating low engagement)

Discussion

The dominance of facial and eye tracking features among the most important signals provides strong empirical support for the effectiveness of webcam-based approaches to engagement measurement. The particular prominence of head pose metrics—especially forward-backward movement and acceleration—suggests that physical orientation toward the screen serves as a key behavioral indicator of engagement. This finding aligns with intuitive expectations: engaged players tend to maintain a more consistent forward orientation, while sudden movements away from the screen may indicate distraction or disengagement.

The strong representation of EmoNet facial embeddings among top features for both high and low engagement highlights the value of deep learning approaches for facial analysis. Although these embeddings lack direct interpretability, their significant contribution to model predictions demonstrates that facial expressions contain rich information about engagement states that may not be easily captured by geometric features alone.

Eye metrics, particularly pupil dilation and blink rate, emerged as important indicators across both engagement states, consistent with what human judges reported using as engagement cues. This finding corroborates previous research linking pupil responses to cognitive load and attentional processes [36], [40]. The bidirectional nature of these signals—appearing in both positive and negative attribution rankings—suggests context-dependent relationships that the model has successfully captured.

The limited representation of EEG features among top contributors aligns with the overall performance hierarchy observed in Table 3.7, where EEG-based approaches (61% accuracy) performed below facial (63%) and flow-based methods (67%). Nevertheless, the appearance of frontal theta activity (AF3.Theta, F8.Theta) among top features for both engagement states is consistent with established literature linking frontal theta oscillations to cognitive engagement and executive function [32], [96].
Interestingly, features showing strong attribution to both high and low engagement (such as `pose_Tz`) highlight the complex, non-linear relationships between physiological signals and engagement states. This complexity likely explains why multimodal approaches demonstrated more balanced detection capabilities across both engagement states compared to single-modality methods.
The attribution patterns observed here provide additional validation for the complementary nature of different sensing modalities in engagement measurement. While facial and eye tracking features appear most prominently among the top contributors, the presence of EEG signals in both positive and negative attribution rankings suggests that neurophysiological data captures unique aspects of engagement that may not be fully reflected in visual cues alone.

3.4.6 Architecture Analysis

To validate our hybrid architecture design, we conducted a systematic comparison against simpler architectures. This analysis aimed to quantify the specific contributions of both the convolutional and transformer components to the overall performance of our engagement estimation framework.
We implemented and compared three architectural variants while controlling for hidden dimension size, training configuration, and input/output interfaces to ensure fair comparison:

- **Hybrid Architecture (Baseline):** The full ConvTransformerEncoder combining convolutional downsampling with transformer-based sequence modeling. This architecture uses a 1D convolutional layer for initial downsampling, followed by ReLU activation, dropout, layer normalization, and a transformer encoder. Global adaptive average pooling produces the final fixed-length representation.

- **Transformer-only (TransformerOnlyEncoder):** Replaces the convolutional layer with a linear projection to the same hidden dimension, followed by dropout and explicit positional encodings. To ensure fair comparison with the convolutional stride in the baseline model, we implemented sequence downsampling by strided sampling of the input sequence. The same transformer encoder configuration and global pooling layer are maintained.

- **Convolution-only (ConvOnlyEncoder):** Replaces the transformer encoder with additional convolutional layers equal to the number of transformer layers. Each additional layer maintains the hidden dimension with residual connections, ReLU activation, and layer normalization to parallel the transformer's structure. The initial downsampling convolutional layer and global pooling remain identical to the hybrid model.

Each architecture variant processed the same input features and was trained with identical optimization parameters, learning schedules, and cross-validation folds. We conducted this analysis on both the multimodal approach (which demonstrated the strongest overall performance among all sensor-based methods) and the webcam-based approach (which offers the best balance between performance and practicality, requiring only standard computer equipment). This dual analysis allowed us to verify whether our architectural findings generalize across different sensing modalities.

Table 3.9: Performance Comparison Across Architecture Variants (Multimodal Approach)

Metric	Low	High	Avg.	ROC_AUC	Accuracy
Multimodal: Hybrid Architecture (ConvTransformer)					
Precision	0.40 ± 0.07	0.76 ± 0.02	0.58 ± 0.03		
Recall	0.41 ± 0.09	0.73 ± 0.06	0.57 ± 0.03	0.59 ± 0.04	0.65 ± 0.03
F1-score	0.38 ± 0.07	0.73 ± 0.04	**0.56 ± 0.03**		
Multimodal: Transformer-only					
Precision	0.31 ± 0.08	0.73 ± 0.03	0.52 ± 0.03		
Recall	0.31 ± 0.10	0.77 ± 0.08	0.54 ± 0.02	0.52 ± 0.06	0.65 ± 0.04
F1-score	0.29 ± 0.08	0.74 ± 0.05	0.51 ± 0.03		
Multimodal: Convolution-only					
Precision	0.32 ± 0.10	0.76 ± 0.02	0.54 ± 0.05		
Recall	0.35 ± 0.13	0.73 ± 0.10	0.54 ± 0.03	0.59 ± 0.04	0.67 ± 0.05
F1-score	0.31 ± 0.10	0.72 ± 0.06	0.52 ± 0.03		

Table 3.10: Performance Comparison Across Architecture Variants (Webcam-based Approach)

Metric	Low	High	Avg.	ROC_AUC	Accuracy
Webcam: Hybrid Architecture (ConvTransformer)					
Precision	0.32 ± 0.07	0.76 ± 0.03	0.54 ± 0.04		
Recall	0.43 ± 0.10	0.69 ± 0.08	0.56 ± 0.03	0.56 ± 0.04	0.63 ± 0.04
F1-score	0.35 ± 0.08	0.71 ± 0.04	0.53 ± 0.03		
Webcam: Transformer-only					
Precision	0.42 ± 0.11	0.72 ± 0.04	0.57 ± 0.05		
Recall	0.23 ± 0.08	0.80 ± 0.08	0.52 ± 0.01	0.56 ± 0.03	0.66 ± 0.05
F1-score	0.22 ± 0.07	0.75 ± 0.05	0.49 ± 0.01		
Webcam: Convolution-only					
Precision	0.37 ± 0.06	0.75 ± 0.02	0.56 ± 0.03		
Recall	0.40 ± 0.09	0.71 ± 0.08	0.55 ± 0.02	0.58 ± 0.04	0.64 ± 0.04
F1-score	0.37 ± 0.06	0.72 ± 0.05	**0.54 ± 0.03**		

Results. For the multimodal approach, the hybrid architecture achieved the highest average F1-score (0.56 ± 0.03) compared to both transformer-only (0.51 ± 0.03) and convolution-only (0.52 ± 0.03) variants. The convolution-only and

hybrid architectures demonstrated equivalent discriminative capability with identical ROC-AUC (0.59 ± 0.04), both outperforming the transformer-only model (0.52 ± 0.06) on this metric. While the convolution-only model showed slightly higher average accuracy (67% ± 0.05 vs. 65% ± 0.03), this difference falls within the standard error margins, suggesting statistically comparable performance when considering confidence intervals.

In the webcam-based approach, a different pattern emerged. The convolution-only architecture achieved the highest average F1-score (0.54 ± 0.03) and ROC-AUC (0.58 ± 0.04), slightly outperforming the hybrid architecture (F1: 0.53 ± 0.03, ROC-AUC: 0.56 ± 0.04). The transformer-only variant, despite achieving nominally higher accuracy (66% ± 0.05), showed the poorest F1-score (0.49 ± 0.01) due to severe class imbalance, with particularly weak low-engagement detection (recall: 0.23 ± 0.08, F1-score: 0.22 ± 0.07).

Importantly, when considering the standard error of the mean (SEM), many of these performance differences fall within overlapping confidence intervals. For instance, the accuracy differences between webcam-based architectures (63% ± 0.04 vs. 66% ± 0.05) cannot be considered statistically significant given their overlapping ranges. This statistical overlap reinforces that our focus should be on consistent patterns across multiple metrics rather than small differences in any single measure.

These results provide important insights into the effectiveness of different architectural components for engagement measurement:

- **Hybrid architecture advantage for multimodal data**: For complex multimodal signals combining EEG, eye tracking, and facial features, the hybrid ConvTransformer architecture achieved the highest F1-score (0.56 ± 0.03), outperforming both single-component alternatives. This suggests that the combination of convolutional feature extraction and transformer sequence modeling provides complementary benefits when processing heterogeneous sensing data.

- **Convolutional strength for webcam data**: For the webcam-based approach, the convolution-only architecture demonstrated the strongest overall performance across metrics (F1: 0.54 ± 0.03, ROC-AUC: 0.58 ± 0.04), suggesting that convolutional processing is particularly effective for facial and head movement features. Importantly, the hybrid approach performed nearly equivalently (F1: 0.53 ± 0.03), with differences falling within the SEM range.

- **Class imbalance handling**: Despite high accuracy in some configurations, the transformer-only architecture exhibited the most severe class imbalance, particularly for webcam data where low-engagement F1-score (0.22 ± 0.07) was dramatically lower than high-engagement (0.75 ± 0.05). This makes it less suitable for real-world applications where detecting both engagement states is important.

- **Statistical significance considerations**: When accounting for SEM, many performance differences between architectures fall within overlapping confidence intervals. This statistical overlap suggests that the major architectural

advantage is not in absolute performance metrics but in consistent, balanced classification across engagement states.

The key finding from this analysis is that combining convolutional and transformer components results in either superior (for multimodal data) or statistically equivalent (for webcam data) F1-scores compared to single-component architectures, while providing more balanced performance across engagement classes. This supports our hybrid architecture design as a robust approach for engagement measurement across different sensing modalities.
Our results align with findings from related domains where convolutional architectures have shown effectiveness for processing physiological[150] signals while transformers excel at capturing sequential dependencies. The hybrid approach leverages these complementary strengths, providing a versatile foundation for engagement measurement in diverse gaming contexts where consistent performance across different engagement states is critical.

3.4.7 Cross-Modal Analysis

To systematically evaluate the complementarity of different sensing modalities, we conducted a pairwise cross-modal analysis. This investigation quantifies how different modality combinations contribute to engagement detection and identifies which pairs offer the strongest synergistic benefits.
We trained binary engagement classifiers for each possible pair of modalities (EEG, Eye, OpenFace, and EmoNet) using our ConvTransformerEncoder architecture with late fusion. The same training configuration and participant-based cross-validation approach described in Section 3.3 was applied to ensure consistent evaluation. For each modality pair, we calculated both F1-scores and ROC-AUC metrics with standard error of the mean (SEM) as uncertainty measure.
As shown in Figures 3.14 and 3.15, the cross-modal analysis reveals several important patterns in modality complementarity. The performance heatmaps demonstrate clear variations in how effectively different modality pairs combine for engagement detection.
The F1-score heatmap (Figure 3.14) shows that combinations involving EmoNet generally achieve the strongest performance. Specifically, the EmoNet+EEG and EmoNet+Eye combinations both achieved the highest F1-score of 0.54±0.04 and 0.54±0.03 respectively. This suggests that facial emotional expressions captured by EmoNet provide complementary information to both neurological (EEG) and visual attention (Eye) signals. The OpenFace+Eye combination showed the weakest performance (0.39±0.04), indicating potential redundancy or conflicting signals between structural facial features and eye movements.
The ROC-AUC heatmap (Figure 3.15) reveals similar patterns with some notable differences. The EmoNet+EEG combination shows the highest ROC-AUC score (0.59±0.04), reinforcing its strength in discriminating engagement states across different thresholds. However, the EmoNet+OpenFace and EmoNet+Eye combinations also perform strongly (0.58±0.03/0.04), suggesting that for threshold-adjustable applications, these combinations offer comparable performance. The Eye+Eye diagonal cell shows the lowest ROC-AUC (0.48±0.04), indicating limitations in using eye-tracking data alone for engagement discrimination.

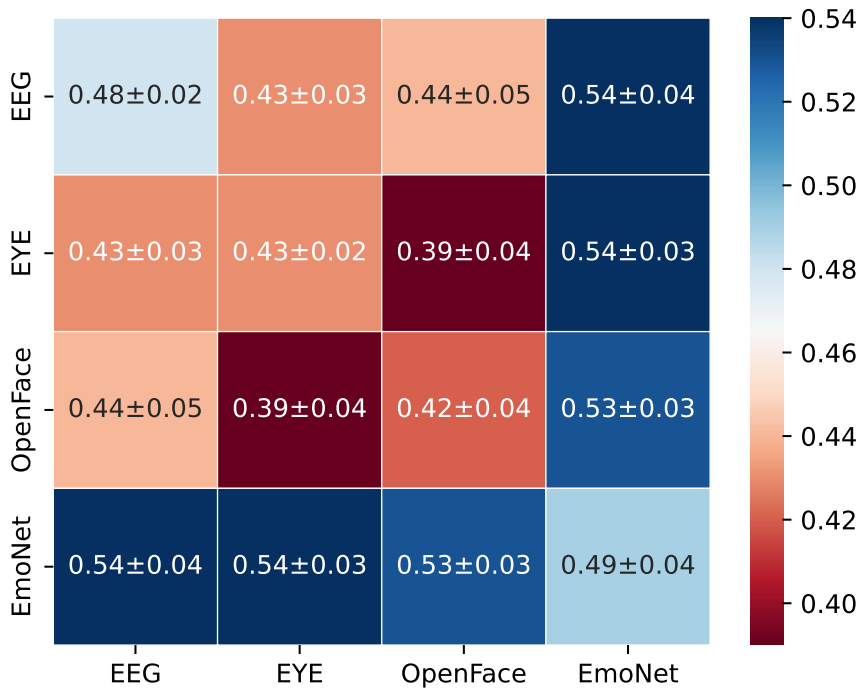


Figure 3.14: F1-score performance heatmap for pairwise modality combinations. Each cell represents the F1-score (with SEM) achieved by combining the modalities from the corresponding row and column. Darker blue indicates higher performance.

These findings provide important guidance for practical multimodal engagement detection systems:

- EmoNet's consistent contribution across multiple combinations confirms the value of deep learning-based facial emotion embeddings for engagement detection, supporting our earlier finding from the ablation study.
- The strong performance of EmoNet+EEG suggests neurological signals capture engagement aspects not apparent in facial expressions, potentially reflecting internal cognitive states without visible manifestations.
- The relatively poor performance of OpenFace+Eye indicates redundancy between these modalities, suggesting both capture similar aspects of visual attention and facial behavior.
- The consistently strong performance of EmoNet-based combinations, compared to the more specialized hardware required for EEG and eye tracking, further supports prioritizing advanced webcam processing for practical applications.

This cross-modal analysis complements our earlier ablation study by specifically quantifying how pairs of modalities perform together rather than measuring performance drops from removing individual components. The results reinforce our finding that facial emotional expressions provide the most valuable engagement signals, while offering new insights into which specific combinations yield optimal performance.

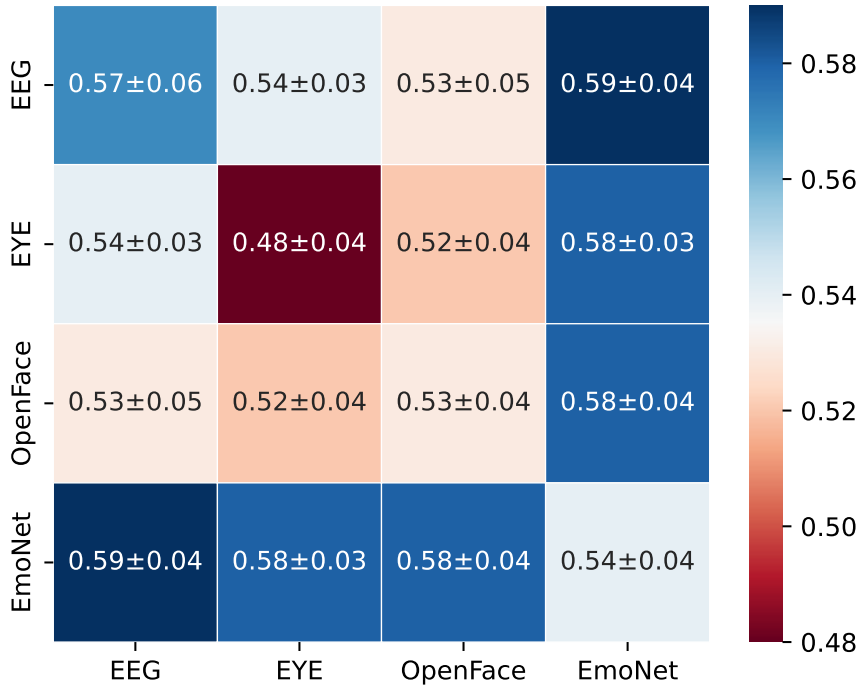


Figure 3.15: ROC-AUC performance heatmap for pairwise modality combinations. Each cell represents the ROC-AUC score (with SEM) achieved by combining the modalities from the corresponding row and column. Darker blue indicates higher performance.

These findings have particular relevance for application scenarios with resource constraints, where deploying the full multimodal system might be impractical. The analysis suggests that combining EmoNet with either EEG or eye tracking provides most of the benefits of the full multimodal approach, offering potential simplification paths for practical implementations while maintaining strong performance.

3.5 Conclusion

This chapter presented the MultiPENG experiment and dataset, a comprehensive multimodal investigation of player engagement in gaming contexts. By collecting synchronized data across multiple sensing modalities while employing a non-intrusive approach to ground truth collection, we provided new insights into the relative effectiveness of different engagement measurement approaches and validated key theoretical models, particularly Flow Theory.

3.5.1 Key Findings and Contributions

Our experimental analysis yielded several significant findings with important implications for engagement measurement in gaming:

Flow Theory Validation. The strong performance of our flow-based model (67% accuracy) using only player skill and game challenge as predictors provides

compelling empirical support for Flow Theory's practical application in engagement estimation. This finding validates Csikszentmihalyi's theoretical framework, which suggests that optimal engagement occurs when skill and challenge are appropriately balanced. The practical implication is that games could potentially optimize engagement by dynamically adjusting difficulty based on player skill, even without complex physiological measurements.

Modality Effectiveness Hierarchy. Our comparative analysis established a clear hierarchy of effectiveness among sensing modalities. The webcam-based approach (63% accuracy) demonstrated strong performance with minimal hardware requirements, followed closely by EEG (61% accuracy) and eye tracking (59% accuracy). When evaluated on a challenging subset of 20 samples, the multimodal approach modestly outperformed human annotators (56% vs. 50% accuracy), while the webcam-based approach underperformed relative to human judges (40-46% accuracy). This mixed comparison with human performance highlights the complex nature of engagement cues and suggests that visual features alone may be insufficient for reliable engagement detection in challenging cases without additional training data or complementary sensing modalities.

Human vs. Computational Assessment. The comparison between human annotators and computational approaches reveals nuanced insights about engagement measurement. The low inter-rater agreement among human judges confirms that engagement lacks consistently interpretable visual indicators, unlike more explicitly manifested states such as happiness or surprise. While the multimodal approach demonstrated an advantage over human annotators, the webcam-based approach's underperformance on the same samples suggests that human observers may leverage contextual cues or implicit knowledge not fully captured by computational analysis of facial expressions and head movements alone. This finding indicates that effective engagement measurement likely requires either multiple sensing modalities or more sophisticated analysis of visual cues with larger training datasets to match or exceed human capabilities.

Multimodal Integration Benefits. While the performance improvement from multimodal integration was modest (65% accuracy), the approach showed more balanced detection capabilities across both high and low engagement states. This suggests that different modalities capture complementary aspects of engagement, which could be valuable in applications requiring nuanced understanding of player states beyond binary classification.

Signal Quality Considerations. Our analysis of signal quality thresholds, particularly for EEG data, demonstrated the critical importance of quality management in physiological signal processing. The optimal 50% quality threshold for EEG signals represents an important methodological contribution for future researchers working with similar data in naturalistic gaming environments.

Statistical Validation. The power analysis revealed complex statistical patterns in our findings. While F1-score metrics achieved high statistical power (> 0.99), substantial effect size fluctuations across different participant samples

(ranging from d=0.64 to d=2.27) indicate that model performance is highly sensitive to specific participants included in the analysis. ROC-AUC results showed more stable effect sizes but achieved sub-optimal power (≈ 0.75), suggesting higher measurement variance. This differential power profile and the observed "cold-start"[147] sensitivity provide important context for understanding both the strengths and limitations of our comparative analysis.

3.5.2 Practical Implications

Our findings have several practical implications for engagement measurement in gaming contexts:

Efficiency-Performance Trade-offs. The relative performance of our approaches suggests different optimal solutions depending on implementation constraints. The flow-based model offers the highest accuracy with minimal computational requirements but depends on having explicit skill and challenge metrics. The webcam-based approach provides strong performance with widely available hardware, making it suitable for large-scale deployment. The multimodal approach offers balanced detection at the cost of increased complexity and hardware requirements.

Detection Bias Considerations. All tested approaches showed substantially higher performance for detecting high engagement states compared to low engagement states. This pattern suggests that engagement manifests more consistently in observable signals during positive engagement experiences, while disengagement may have more varied expressions that are harder to model. This detection bias should be considered when deploying these systems, especially in applications where identifying disengagement is particularly important.

Human vs. Computational Assessment. The substantial performance gap between human annotators and computational approaches highlights the subtle nature of engagement cues. Unlike more explicitly manifested states like happiness or surprise, engagement appears to lack consistently interpretable visual indicators for human observers. This finding underscores the value of computational approaches that can identify patterns not readily apparent to human observers.

3.5.3 Limitations and Future Work

Several limitations of our study should be acknowledged:

Dataset Characteristics. Our evaluation used binary engagement classification (low/high) rather than the original 5-point scale to mitigate class imbalance and ensure sufficient samples per class. While this approach provides clear comparative insights, it simplifies the full spectrum of engagement experiences. Future work with larger datasets may explore more granular classification or regression approaches.

Participant Sensitivity. The significant fluctuations in effect sizes observed during our power analysis (d ranging from 0.64 to 2.27 for F1-score) highlight a "cold-start problem" where model performance is highly sensitive to which specific participants are included in the dataset. This sensitivity suggests that engagement manifestation varies substantially across individuals, potentially limiting the generalizability of our models to new populations without additional calibration or personalization strategies.

Challenge Measurement Limitations. The flow-based model's performance suggests potential for even greater accuracy if more sophisticated skill and challenge metrics were available. In our implementation, challenge was represented by predefined difficulty levels, which may not fully capture the dynamic challenges experienced by players with varying skill levels. Similarly, skill was represented by self-reported experience levels, which provide only an approximate measure of actual player capability.

Contextual Factors. Our evaluation focused on statistical performance of different measurement approaches and did not fully account for variations in effectiveness across different gaming genres, player demographics, or gameplay contexts. The relative performance of these approaches may vary in different settings or with different player populations.

3.5.4 Motivating Next Chapter

The findings from this exploratory work establish a foundation for addressing our key research directions identified in Chapter 2. The effectiveness of the flow-based model, in particular, suggests promising avenues for developing non-intrusive, real-time engagement measurement in complex gaming environments. Rather than requiring players to report their skill levels or relying on predetermined difficulty settings, the next chapter focuses on dynamically estimating both skill and challenge from readily available game telemetry data. This approach enables the practical application of Flow Theory-based engagement estimation in a wide range of gaming contexts without disrupting the player experience.
By identifying that Flow Theory provides a robust framework for engagement estimation, we can now focus on developing computationally efficient methods for extracting the necessary skill and challenge metrics from gameplay data. This shift from measuring the manifestation of engagement in the player (through physiological and behavioral signals) to examining how it manifests in gameplay itself represents a key progression in our research agenda, addressing the need for non-intrusive, real-time measurement methods that can be implemented in complex gaming environments.

Chapter 4

Real-Time Player Engagement Measurement Using Non-Intrusive Game Telemetry

This chapter makes verbatim reuse or rephrasing of the material in the following paper, with permission [79]:
Ammar Rashed, Shervin Shirmohammadi, and Mohamed Hefeeda, "Real-Time Player Engagement Measurement Using Non-Intrusive Game Telemetry", *IEEE Open Journal of Instrumentation and Measurement*, Volume 4, 2025, Article Sequence Number 2500116, 16 pages.
The dataset used in this chapter is publicly available on Kaggle (10.34740/KAGGLE/DS/7099170).

4.1 Introduction

Building on the conceptual foundations established in Chapter 2, this chapter presents a novel framework for real-time, non-intrusive measurement of player engagement. As previously discussed, measuring and monitoring player engagement has become crucial for game developers and researchers alike [151], with significant implications for both the entertainment value and commercial success of video games. While Chapter 2 provided a comprehensive review of engagement concepts and measurement approaches, this chapter focuses specifically on addressing the practical challenge of implementing real-time, non-intrusive instrumentation methods that can capture engagement's multidimensional nature without disrupting the player experience [43], [44].
This chapter addresses the specific research gap identified in Chapter 2 regarding non-intrusive, real-time engagement measurement methods that can be implemented at scale without disrupting the player experience. The practical importance of such methods extends beyond academic interest to key applications in entertainment, education [152], and business domains such as churn prediction [153], [154] and data-driven game design improvements [134], [155], [156].
Precise measurement of player engagement in real-time faces multiple challenges:

1. Player engagement is a complex, multi-dimensional construct requiring simultaneous monitoring of cognitive, behavioral, and emotional signals

2. Engagement manifests both as an instantaneous measurable state and as an evolving process over time, complicating the development of unified measurement models

3. Players' diverse preferences and gaming experiences necessitate adaptive measurement approaches across different game contexts

Traditional methods for measuring player engagement rely on two main approaches: post-game surveys and physiological data collection [5], [6]. As explored in Chapter 2, post-game surveys, such as the Game Experience Questionnaire (GEQ) [3], provide comprehensive measurement data and are straightforward to implement. However, they suffer from recall bias due to the time gap between gameplay and reporting [4], and cannot capture temporal fluctuations in engagement signals [157]. Conversely, physiological measurements, such as those explored in the MultiPENG study (Chapter 3), offer continuous, objective signal collection during gameplay. While this approach provides high-resolution temporal data, it requires specialized sensing equipment, creates artificial measurement conditions, and typically involves complex signal processing that prevents real-time engagement detection.

4.1.1 Flow Theory-Based Approach

Our measurement approach is grounded in Flow Theory [7], which, as established in Chapter 2, posits that optimal engagement occurs when a player's skill matches the game's challenge. As discussed in the background chapter, Flow Theory has been implemented through various models including the Quadrant Model [76], Experience Fluctuation Model [77], and Flow Channel Model, each offering valuable perspectives on how skill-challenge relationships influence player experience.
The MultiPENG study in Chapter 3 provided strong empirical validation for Flow Theory as a foundation for engagement estimation. The Flow-based model using only skill and challenge measures achieved 66±2% accuracy (F1-score: 0.59±0.03) in predicting engagement states, outperforming sophisticated multimodal approaches that combined EEG, eye tracking, and facial features (65±3% accuracy, F1-score: 0.56±0.03). This finding is particularly significant considering that human observers achieved only 50±3% accuracy when attempting to visually assess engagement from webcam footage. Moreover, the Flow-based model demonstrated better balanced performance across both high engagement (precision: 0.77±0.03) and low engagement states (precision: 0.43±0.06) compared to other modalities. This superior performance, achieved with just two variables, suggests that Flow Theory provides a remarkably efficient framework for engagement estimation in gaming contexts. The study validated the effectiveness of this approach in a controlled setting, but skill and challenge remain difficult to measure directly and non-invasively during actual gameplay.

4.1.2 Limitations of Current Approaches

As detailed in Chapter 2, recent engagement estimation methodologies have explored diverse approaches including facial expression analysis [40], physiological measurements [87], and multimodal techniques. However, our analysis of

these approaches in the MultiPENG study revealed significant practical limitations. Webcam-based models showed poor precision (0.32±0.07) for low engagement states despite achieving 63±4% overall accuracy. EEG-based models exhibited similar imbalances with 61±5% accuracy, while eye tracking features achieved only 59±6% accuracy. Even combined multimodal approaches reached just 65±3% accuracy with continued struggles in low engagement detection (precision: 0.40±0.07).

These findings from the MultiPENG study directly inform our current approach. Rather than collecting multiple physiological signals requiring specialized equipment, we focus on deriving the two most predictive variables—skill and challenge—from readily available game telemetry data. This approach builds on Flow Theory's demonstrated effectiveness while addressing the practical limitations of sensor-based data collection.

The comprehensive review in Chapter 2 noted how game telemetry has increasingly been used to understand player behavior and experience. Of particular relevance to our approach are works by Melhart et al. [116], who used in-game events to model player experience, and Reguera et al. [39], who explored gameplay session data as a proxy for engagement. Especially noteworthy is the study by Melhart et al. [115] that demonstrated the power of gameplay features in predicting viewer engagement on Twitch streams specifically for PUBG, achieving prediction accuracies of up to 80% by analyzing the relationship between in-game events and viewer chat frequency. This promising result with PUBG telemetry data directly influenced our choice of experimental domain.

4.1.3 Skill and Challenge Estimation

Previous approaches to measuring skill and challenge, as reviewed in Chapter 2, have typically relied on predetermined difficulty settings or self-reported skill levels, which prove impractical for real-time applications in complex gaming environments. Notable examples include Aponte et al. [158], who used reinforcement learning to train virtual agents and measure challenge based on agent performance, and Wheat et al. [159], who analyzed level characteristics in 2D games to model challenge. For skill estimation, Diah et al. [160] used heuristics like the number of enemies defeated in MOBA games.

While these methods provide valuable insights, they often lack generalizability across different game genres and struggle to capture the dynamic nature of multiplayer online games. Our work builds on these foundations while addressing their limitations through a telemetry-based approach that measures skill through relative competitive performance and challenge through immediate survival threats, creating a more universal methodology that functions across various competitive gaming environments.

As reviewed in Chapter 2, much of the instrumentation and measurement literature focuses on user engagement in medical applications, such as rehabilitation [161], ADHD detection systems [162], medical device risk assessment [163], and dementia care [60]. However, few approaches have attempted to leverage game telemetry data directly for engagement measurement. An exception is [164], which uses physiological measures to assess driver engagement. While [162] and [60] incorporate games in their methodologies, none of these works utilize game telemetry

data to measure engagement in real time, which is a main novelty of our proposed framework.

4.1.4 Our Telemetry-Based Framework

To address the limitations of existing approaches, we propose a non-intrusive framework for real-time measurement of player engagement using game telemetry signals. Modern games routinely collect comprehensive telemetry data about player actions, performance metrics, and game states, providing an accessible and scalable measurement source.

Our telemetry-based framework addresses these limitations by measuring skill through relative competitive performance and challenge through immediate survival threats, creating a more universal approach that functions across various competitive gaming environments. We use easier-to-measure telemetry signals including combat statistics, movement patterns, resource management, and general match states, to then indirectly measure skill and challenge. To do so, we use Graph Convolutional Networks (GCN) that detect complex player interactions and spatial relationships, coupled with Transformer networks that process temporal sequences of game states. This hybrid architecture produces two proxy metrics that quantify skill and challenge, as described next.

The first proxy metric is the player's ranking in the match, which indicates their skill level. Ranking is usually determined at match completion, but our framework detects likely match outcomes in real-time based on ongoing performance signals. Higher ranking indicates the player outperforming others, suggesting higher skill levels. We consider skill to be a continuous ordinal quantity normalized between 0 (lowest skill) and 1 (highest skill).

The second proxy metric is the total damage sustained from both enemy attacks and environmental hazards per game phase, which quantifies the challenge level. Higher damage indicates relatively more difficult game conditions during that phase. We consider challenge to be a continuous, positive, and unbounded ordinal quantity.

Finally, the measured skill and challenge are fed to an engagement measurement module - a binary classifier which uses an established baseline from player survey data to classify engagement as an ordinal quantity of either 0 (low engagement) or 1 (high engagement). The entire measurement pipeline operates in real-time, providing engagement estimates that precede actual gameplay outcomes by variable time intervals, depending on when the player is eliminated or the phase ends.

Our validation approach addresses the challenges in obtaining reliable engagement annotations through a hierarchical framework. First, we establish an empirical baseline by analyzing self-reported skill-challenge-engagement relationships. Then, we validate our telemetry-based proxies against these self-reports to ensure alignment with player perceptions. Finally, we evaluate our real-time estimation system end-to-end by comparing its predictions against post-game survey responses. This enables systematic validation from raw telemetry data to final engagement predictions.

Building on the findings from the MultiPENG study in Chapter 3, which demonstrated the effectiveness of Flow Theory-based metrics in predicting engagement, this chapter presents a practical realization of these insights through a telemetry-

based framework. The primary contributions of this work can be summarized as:

- An instrumentation methodology for real-time measurement of engagement based on flow theory, transforming standard game telemetry signals into continuous skill and challenge measurements without gameplay interruption.
- A hybrid signal processing architecture combining GCN for player interactions with Transformers for temporal sequences, demonstrating superior measurement performance over single-architecture alternatives.
- A high-resolution engagement measurement methodology using survey-calibrated baselines, enabling detection of significant variations within inherently engaging game contexts.
- A practical realization of the measurement framework in PUBG, demonstrating its viability for complex multiplayer environments with diverse gameplay mechanics and player interactions.
- Cross-domain validation of the model, as is and without transfer learning, with FIFA'23, a sports game, and Street Fighter V, a fighting game, demonstrating that the approach is genre-agnostic, applicable to a wide variety of game types beyond combat-focused games, including sports games, racing games, strategy games, and more.

The real-time engagement metrics provided by our framework offer several practical applications for game developers. First, they enable dynamic difficulty adjustment, where the game can automatically modify challenge levels based on detected engagement states, preventing player frustration or boredom [8], [9]. Second, they facilitate targeted content delivery, allowing developers to introduce new gameplay elements or narrative sequences precisely when engagement begins to decline [10]. Third, they support personalized matchmaking systems that can maintain optimal skill-challenge balances across different player segments [11], [12]. Fourth, they enable intelligent resource allocation in cloud gaming environments, where streaming quality and latency significantly impact user engagement [13].
By identifying highly engaging gameplay moments, cloud gaming providers can dynamically allocate more bandwidth and processing resources to maintain quality during critical periods, similar to how video providers optimize streaming quality to maximize user engagement [165]. Beyond individual player optimization, our metrics can reveal engagement patterns across different game features, scenarios, play sessions, and platforms, enabling developers to identify which specific game elements consistently drive or diminish engagement. The measurement approach we propose provides several key features that address the research gaps identified in Chapter 2: temporal granularity (detecting engagement fluctuations within individual matches), contextual awareness (understanding engagement in relation to specific gameplay contexts), scalability (processing data from thousands of concurrent players), and actionability (producing metrics that directly inform design decisions). Unlike post-hoc analysis, these real-time capabilities enable proactive interventions that can significantly impact player retention and satisfaction.

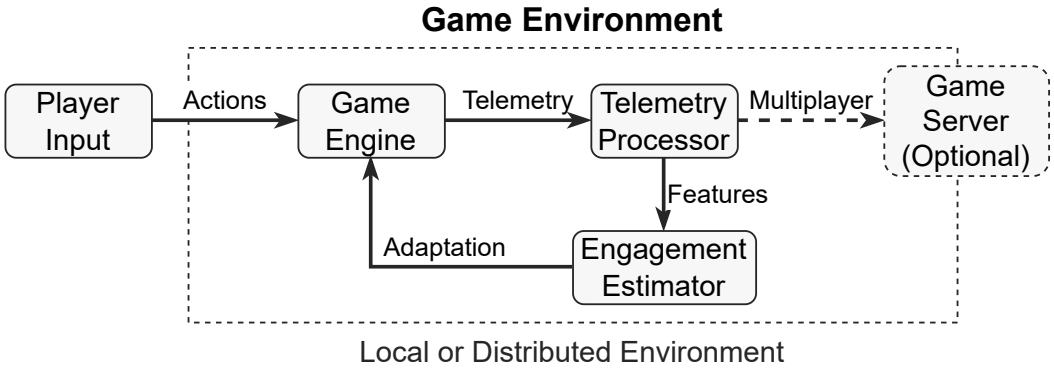


Figure 4.1: The considered model for gaming systems.

The rest of this chapter is organized as follows. Section 4.2 describes the proposed framework and problem formulation. In Section 4.3 we present the proposed solution, while in Section 4.4 we discuss the PUBG case study and analyze the performance evaluation results. The chapter is concluded in Section 4.5.

4.2 System Model and Problem Definition

In this section, we specify the considered model for gaming systems and formally define the player engagement problem.

4.2.1 System Model

Our engagement estimation framework operates within a general gaming environment, as illustrated in Figure 4.1. The Game Server authenticates players and matches them in games. Players send actions to the Game Engine, which renders the game and maintains the game state. The Game Engine also logs telemetry events containing details about shots fired, weapons used, locations of the attacker and victim, etc. The Telemetry Processor aggregates these raw events into meaningful features. Consider, for example, a combat scenario where a player attacks an opponent. The Telemetry Processor converts combat events into metrics like damage dealt and accuracy rates and movement events into distance traveled. It also converts inventory events into resource utilization patterns. These processed features capture both player skill (through combat performance and resource management) and challenge levels (through damage taken and threat proximity). Finally, the Engagement Estimator analyzes the high-level telemetry features to measure player engagement. It can be integrated directly within the Game Engine for immediate state updates or implemented as an external module when handling complex multiplayer scenarios requiring additional processing capacity.
To illustrate with a practical example from our PUBG case study: when a player engages in combat, the Game Engine logs raw events such as "Player A fired a weapon," "Player A hit Player B," and "Player B took X damage." The Telemetry Processor aggregates these into meaningful features including accuracy (hits/shots), damage per minute, and combat efficiency (damage dealt/damage taken). Simultaneously, movement events like "Player A moved to position (x,y,z)" are

transformed into metrics such as distance traveled, rotation frequency, and positioning relative to safe zones. Consider a specific in-game scenario where a player engages an opponent at medium range using an assault rifle. The Telemetry Processor would capture combat performance (e.g., 60% accuracy, 90 damage dealt), positioning context (e.g., partial cover utilization, high ground advantage), resource management (e.g., ammunition consumption rate, healing item usage), and threat assessment (e.g., proximity to other teams, position relative to play zone boundary). These processed features would then feed into the Engagement Estimator to determine the player's current skill expression and challenge level during this combat interaction.

This flexible architecture supports various deployment scenarios, from single-player games with local processing to distributed multiplayer environments. In online multiplayer games, clients connect to game servers that aggregate player interactions and state updates, allowing the engagement estimator to operate at the server level to account for inter-player dynamics. The system assumes reliable telemetry data collection and low-latency processing capabilities to enable real-time engagement estimation.

To capture meaningful interactions in games, we represent the game state at time t as a dynamic graph $G(t) = (V(t), A(t))$, where $A(t)$ represents the adjacency structure. The vertices $V(t)$ represent game entities (players, non-player or AI characters, interactive objects, or environmental elements), while the adjacency structure encodes relevant relationships between these entities. These relationships can be defined flexibly based on game-specific interaction metrics such as spatial proximity, direct interaction, strategic relevance, or causal relationships.

For each vertex $v \in V(t)$, we maintain a feature vector $\mathbf{x}_v(t)$ that captures its current state. This graph representation is versatile and can model various game scenarios: competitive or cooperative interactions between human players, interactions with AI-controlled enemies or environmental hazards, relationships between team members, or interactions between a player and game-generated entities in single-player games.

This system model is adaptable to various gaming scenarios. In single-player games, the graph representation simplifies as interactions occur primarily between the player and game-generated entities (environment, AI characters, objectives). Local multiplayer scenarios can be modeled with stronger emphasis on direct player-to-player interactions, often with richer adjacency structures reflecting physical proximity and shared interfaces. In online multiplayer contexts, as demonstrated in our PUBG case study, the model captures complex player-to-player interactions across potentially large networks, team-based dynamics, and player-environment interactions. The graph structure can flexibly represent competitive relationships (as negative edges or repulsive forces), cooperative alliances (as positive edges or attractive forces), or neutral interactions based on proximity or shared objectives. This flexibility enables our engagement estimation framework to accommodate diverse gaming modalities while maintaining a consistent mathematical formulation and measurement approach.

4.2.2 Problem Definition

Given a multiplayer online game environment, as described in the above system model, we consider the problem of estimating player engagement in real time. Formally, at any time t during gameplay, for each player p, we aim to estimate their current engagement level $E_p(t)$ based on the historical telemetry data available up to time t, denoted as $\mathcal{H}_{\text{tele}}(t)$. This telemetry includes player actions, game states, and interaction patterns captured through standard game logs.

To concretize this problem definition, consider a player in a PUBG match at time t = 5 minutes into the game. The historical telemetry data $\mathcal{H}_{\text{tele}}(t)$ would include all player actions and game states up to that moment, such as the player's weapon acquisition sequence, early-game positioning decisions, initial resource gathering efficiency, and any early combat encounters. Our framework aims to estimate their current engagement level $E_p(t)$ based on these observable telemetry patterns before key gameplay outcomes materialize. For instance, the framework might detect declining engagement when a player's movement patterns become erratic after failing to find adequate equipment, allowing for potential interventions (such as nearby loot spawns) before the player becomes fully disengaged. Importantly, this estimation occurs without requiring any explicit feedback from the player, relying solely on behavioral signals captured through standard game logs.

This formulation advances beyond traditional engagement estimation approaches by emphasizing predictive capabilities - estimating engagement before critical gameplay moments materialize, rather than retroactively analyzing completed sessions. It also acknowledges the temporal nature of engagement, treating it as a dynamic measure that evolves throughout gameplay rather than a static post-game metric. While this real-time constraint introduces additional complexity, it enables practical applications in dynamic game adaptation.

In addition, our formulation does not require any intrusive physiological measurements or post-game questionnaires. It only utilizes standard telemetry data, which ensures broader applicability across existing game infrastructures and maintains non-intrusive monitoring of player experiences.

Unlike the multimodal approach explored in the MultiPENG study (Chapter 3), which required specialized equipment to capture physiological signals, our telemetry-based approach can be deployed at scale without additional hardware requirements. This addresses a key limitation identified in previous engagement measurement approaches while building on the insight that Flow Theory-based metrics provided strong predictive performance in our controlled experiments.

4.3 Proposed Solution

This section first specifies the design goals of the proposed framework and presents an overview of how it functions. It then describes the details of each component.

4.3.1 Design Goals and Solution Overview

The proposed framework is designed to achieve the following goals.

- **Relative Engagement Scale**: By comparing current engagement levels to an established baseline $\bar{E}$, we enable the detection of meaningful engagement

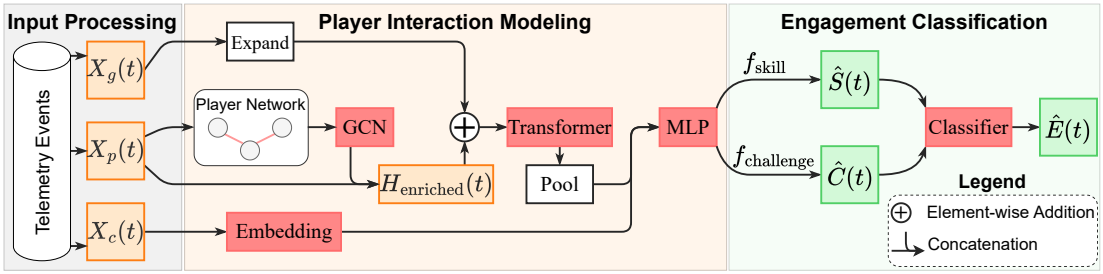


Figure 4.2: Overview of the proposed engagement estimation framework. Game telemetry data flows through the GCN and Transformer networks to predict the skill and challenge metrics, which are used to estimate engagement $\hat{E}(t)$.

variations while maintaining computational efficiency.

- **Flow Theory-Based Quantification**: Engagement is quantified through the relationship between player skill and game challenge, requiring only standard telemetry data while maintaining theoretical grounding.
- **Non-Intrusive Instrumentation**: The framework exclusively utilizes game telemetry data available through standard logging systems, enabling scalable deployment without additional hardware requirements.
- **Hybrid Measurement Approach**: The framework combines theoretical foundations with machine learning-assisted measurement techniques, enabling both interpretable and accurate engagement measurement while maintaining flexibility for different game genres and contexts.

While telemetry data may occasionally be unavailable due to disabled logging systems, incomplete instrumentation, or privacy-focused game configurations, this scenario is exceptionally rare in modern commercial games. Since telemetry fundamentally represents any recordable player interaction—from basic input events to gameplay state changes—virtually all interactive games generate inherent telemetry data by design. The concern shifts from whether telemetry data exists (as player interactions are core to gaming) to whether logging systems are actively capturing and storing this naturally occurring interaction data. Even minimalist games produce trackable events through fundamental mechanics like player input, state transitions, and session boundaries, making telemetry absence primarily a matter of deliberate logging disablement rather than data non-existence.
Figure 4.2 presents an overview of the proposed framework for real-time player engagement estimation. The framework comprises three main components: input processing, player interaction modeling, and engagement classification. At a high level, telemetry data is first processed to extract representative features characterizing both player behavior and game state. A dynamic player interaction network is constructed, where edges can represent various types of relationships such as spatial proximity, direct interactions, team affiliations, or shared object interactions. This network, along with the processed features, is then enriched through a hybrid neural architecture combining GCNs for modeling player interactions and Transformer networks for capturing temporal patterns. The enriched features are used to measure player skill and game challenge levels, which are then mapped to binary engagement states using a Random Forest classifier.

We provide the details of each component in the following subsections. First, we describe the input processing module that handles various types of telemetry features and constructs the player interaction network. Then, we elaborate on the player interaction modeling component that enriches these features through graph-based and temporal processing. Finally, we present the engagement classification module that estimates skill and challenge levels to determine player engagement states.

4.3.2 Architecture Justification

Our hybrid GCN-Transformer architecture addresses two fundamental challenges in player engagement modeling: capturing complex spatial interactions between players and tracking temporal dynamics of gameplay. The selection of this particular architecture is motivated by both theoretical considerations and empirical findings.

Graph Convolutional Networks (GCNs) are particularly well-suited for modeling complex player interactions in gaming environments. Recent work by Xenopoulos et al. [166] demonstrated the effectiveness of graph-based representations for game state modeling in both traditional sports and esports contexts. GCNs preserve permutation invariance, maintain spatial relationships, and allow flexible player interaction weights—crucial advantages when modeling battle royale games like PUBG where player positions and team interactions significantly impact performance and engagement. Additionally, Stöckl et al. [167] showed that GCNs can effectively model interaction patterns without relying on predefined role assignments, making them adaptable to the fluid nature of battle royale gameplay where player roles evolve dynamically throughout a match.

The Transformer component of our architecture addresses the temporal aspects of engagement by capturing sequential dependencies in gameplay. Simpson et al. [168] demonstrated that Transformer-based models excel at learning the sequential language of gameplay events, enabling predictions that account for both immediate and long-range temporal dependencies. This capability is essential for modeling engagement as a dynamic process that evolves throughout a match rather than a static state. We empirically validated this architectural choice by comparing against simpler sequential models including RNN, GRU, and LSTM, which consistently provided lower performance. The superior results of the Transformer architecture can be attributed to its multi-head attention mechanism, which enables selective focus on relevant gameplay events across different time scales and contexts, rather than the sequential bottleneck inherent in recurrent architectures that can lose important temporal information over long sequences.

The hybrid approach combining these architectures was chosen after extensive experimentation with alternative configurations. For instance, we tested a simplified architecture that fed all three categories of features (player, game state, and categorical) directly into the GCN before passing the output to the Transformer. While this resulted in a more streamlined design, it led to significantly degraded performance. This performance difference can be attributed to the hybrid architecture's ability to process different types of information through specialized pathways:

1. The GCN focuses on player-to-player interactions while preserving individ-

ual player characteristics

2. The Transformer processes temporal sequences with direct access to both interaction-enriched player features and global game state information

3. The categorical embeddings maintain their semantic integrity by bypassing the graph convolution process

This separation of concerns allows each component to specialize in processing its respective type of information before integration in the final layers. Moreover, the hybrid architecture's superior performance aligns with findings in multimodal learning research, where specialized pathways for different data modalities often outperform unified processing approaches [169].
Our ablation studies (see Section 4.4.2) further validate this architectural choice, showing that both the GCN and Transformer components make substantial contributions to the framework's overall performance. The hybrid architecture (accuracy: 0.73±0.14, ROC-AUC: 0.83±0.17) significantly outperforms the Transformer-only variant (accuracy: 0.67±0.19), demonstrating the value of modeling player interactions through graph convolution.

4.3.3 Input Processing

Building upon the telemetry processing system described in Section 4.2.1, our framework transforms the real-time stream of raw telemetry events into three types of high-level features. These features are computed and sampled at fixed intervals (e.g., every 10 seconds) to create consistent temporal snapshots of the gameplay state.

- **Player Features** ($X_p(t) \in \mathbb{R}^{T \times N \times d_p}$): These features characterize individual player performance and behavior patterns. For each player, we aggregate telemetry events into meaningful metrics capturing combat performance (e.g., accuracy, damage dealt), mobility (e.g., distance traveled, position changes), resource utilization (e.g., item usage, inventory management), and spatial awareness (e.g., proximity to threats, zone positioning). Here, T represents the sequence length, N is the number of players, and d_p is the feature dimension. These metrics serve as proxies for player skill and adaptability.

- **Game State Features** ($X_g(t) \in \mathbb{R}^{T \times d_g}$): These features capture the evolving match context and environmental conditions that affect all players. By tracking match progression metrics such as elapsed time, remaining players, and zone states, we can contextualize individual player behaviors and better estimate the current challenge level. The dimension d_g represents the game state feature space.

- **Categorical Features** ($X_c(t)$): Discrete contextual information such as game mode and phase information is encoded through embedding layers. These features provide essential context for interpreting player behaviors and performance metrics, as similar actions may have different implications across different game modes or phases.

A comprehensive list of all features used in our implementation, including their definitions, data types, and value ranges, is provided in Appendix B. This includes detailed descriptions of the player features, game state features, and categorical features utilized in our PUBG case study.

4.3.4 Player Interaction Modeling

The player interaction modeling component processes the input features through a hybrid neural architecture designed to capture both spatial and temporal relationships in gameplay. Following the graph-based game state representation defined in subsection 4.2.1, we first construct a dynamic player network where nodes represent players and edges represent their relationships. The edge weights w_{ij} between players i and j can encode various types of interactions such as spatial proximity, direct combat engagement, or team-based cooperation.
This player network is processed through a GCN consisting of multiple layers with decreasing dimensionality to learn compact representations that capture the structural properties of player interactions. The GCN architecture employs skip connections between layers to preserve individual player features while learning interaction-based representations. The GCN outputs are then combined with the original player features through concatenation to create enriched representations:

$$\mathbf{H}_{\text{enriched}}(t) = [\text{GCN}(X_p(t)); X_p(t)] \tag{4.1}$$

To capture temporal dependencies, we employ a multi-layer Transformer encoder with multiple attention heads. Before processing, the game state features $X_g(t)$ are expanded along the player dimension to enable element-wise operations with the player-specific features in $\mathbf{H}_{\text{enriched}}(t)$. The Transformer processes these combined features through self-attention mechanisms, enabling the model to identify relevant temporal patterns and long-range dependencies in player behavior. The multi-head attention architecture allows the model to capture different aspects of temporal relationships simultaneously:

$$\mathbf{H}_{\text{temp}}(t) = \text{Transformer}(\mathbf{H}_{\text{enriched}}(t), X_g(t)) \tag{4.2}$$

$$\mathbf{H}_{\text{pool}}(t) = \text{Pool}(\mathbf{H}_{\text{temp}}(t)) \tag{4.3}$$

The categorical features are processed through embedding layers that map each discrete feature to a lower-dimensional dense representation $\mathbf{H}_{\text{cat}}(t) = \text{Embed}(X_c(t))$. All processed features are then combined through a multi-layer perceptron (MLP) with specialized output heads for skill and challenge measurement:

$$\mathbf{Z}(t) = \text{MLP}([\mathbf{H}_{\text{pool}}(t); \mathbf{H}_{\text{cat}}(t)]) \tag{4.4}$$

This hierarchical processing enables our framework to capture complex player interactions at multiple scales while maintaining the temporal context necessary for engagement measurement.
The specific hyperparameters used in our implementation, including the number of GCN layers, Transformer architecture details, learning rates, and training procedures, are provided in Appendix C. This information is included to ensure reproducibility and to provide guidance for adapting the framework to other gaming contexts.

4.3.5 Engagement Classification

The final component produces engagement measures through a two-step process that explicitly models the relationship between player skill and game challenge. First, from the processed features $\mathbf{Z}(t)$, we measure skill and challenge levels through separate prediction heads:

$$\hat{S}(t) = f_{\text{skill}}(\mathbf{Z}(t)) \tag{4.5}$$

$$\hat{C}(t) = f_{\text{challenge}}(\mathbf{Z}(t)) \tag{4.6}$$

The activation functions for these measures are chosen based on the nature of the underlying skill and challenge proxies. For example, when using normalized ranking as a skill proxy, we employ a sigmoid activation for f_{skill} to bound the output between 0 and 1. In contrast, when using metrics like damage received as challenge proxies, we utilize Rectified Linear Unit (ReLU) activation for $f_{\text{challenge}}$ to handle unbounded positive values. This choice of activation functions can and should be adapted based on the specific proxies used in different game contexts.
The measured skill and challenge levels are then mapped to binary engagement states through a Random Forest classifier:

$$\hat{E}(t) = f_{\text{classify}}(\hat{S}(t), \hat{C}(t)) \tag{4.7}$$

where engagement is defined relative to a baseline $\bar{E}$:

$$E_{\text{binary}}(t) = \begin{cases} 1 & \text{if } E(t) > \bar{E} \text{ (High Engagement)} \\ 0 & \text{if } E(t) \leq \bar{E} \text{ (Low Engagement)} \end{cases} \tag{4.8}$$

The baseline $\bar{E}$ is established through a one-time calibration process using self-reported engagement levels from player surveys. This calibration is crucial for inherently engaging game genres, such as competitive multiplayer games, where most players maintain some baseline level of engagement. In such contexts, the baseline helps distinguish subtle variations in engagement levels rather than merely detecting obvious disengagement. While our implementation uses survey responses for baseline calibration, alternative approaches such as expert annotations or behavioral indicators could be used depending on the available data and specific game context.
The choice of Random Forest for the final classification aligns with the non-linear nature of the skill-challenge relationship in Flow Theory. It can capture complex decision boundaries between engagement states while providing interpretable feature importance scores that help validate the relative impact of skill and challenge on engagement predictions.
Building on the insights from the MultiPENG study in Chapter 3, which demonstrated the effectiveness of Flow Theory-based engagement prediction, our approach extends this concept to dynamic, real-time gameplay environments. Rather than relying on self-reported skill levels and predetermined difficulty settings as in the controlled MultiPENG experiments, our framework derives these measures directly from observable gameplay behaviors, enabling scalable deployment in complex gaming environments.

4.4 Evaluation

We evaluate our engagement estimation framework through a systematic validation process, focusing both on component-level performance and end-to-end effectiveness. Our evaluation employs PlayerUnknown's Battlegrounds (PUBG) as a case study, leveraging its rich telemetry data and diverse gameplay mechanics to thoroughly assess our framework's capabilities in a real-world setting. Throughout this evaluation section, all reported uncertainties (±) represent the standard error of the mean (SEM) of the corresponding metric, similar to Section 3.4.

4.4.1 Experimental Setup

Framework Implementation

We implement our framework for PUBG, a battle royale game where approximately 100 players compete in teams across large maps, starting with no equipment and scavenging for resources while avoiding elimination. The game naturally segments into distinct phases as the playable area progressively shrinks, with each phase typically lasting 2-3 minutes. For skill and challenge quantification, we define:

$$S_p(t) = \frac{\text{number of players eliminated before p}}{\text{total number of players - 1}} \tag{4.9}$$

$$C_p(t) = \text{total damage taken by player p in phase t} \tag{4.10}$$

Our implementation samples telemetry data at 10-second intervals. The feature dimensions are:

- $X_p(t) \in \mathbb{R}^{T \times N \times 35}$ for player features
- $X_g(t) \in \mathbb{R}^{T \times 4}$ for game state features
- $X_c(t)$ includes map identifier, team size, and phase index

The player interaction graph $G(t)$ is constructed with edge weights:

$$w_{ij} = \max(0, 1 - \frac{d_{ij}}{d_{max}}) \cdot [(1 - \sigma(\theta)) \cdot \mathbb{1}_{enemy} + \sigma(\theta) \cdot \mathbb{1}_{teammate}] \tag{4.11}$$

where d_{ij} is the Euclidean distance between players (capped at $d_{max} = 100$ meters), θ is a learnable team weight parameter, and $\mathbb{1}_{enemy}$, $\mathbb{1}_{teammate}$ are binary indicators for enemy/teammate relationships.

There's potential confusion about whether skill constitutes "one label per match" since final ranking remains constant throughout a game. However, this overlooks that the set of alive players changes at each game phase, creating dynamic ranking contexts where the same final ranking translates to different relative skill assessments depending on the current player pool. The sequential model is essential because it must learn to predict how current gameplay patterns will translate to final outcomes given the evolving competitive landscape—a player ranked 5th overall represents different skill levels when competing against 50 players versus 10 remaining players, and the model must capture these contextual nuances through temporal gameplay sequences.

A potential concern with using total damage taken as a challenge proxy is that experienced players might deliberately engage in high-risk, high-damage scenarios as part of aggressive strategies rather than due to insufficient skill. However, this criticism applies to virtually any telemetry-based proxy—player behavior inherently contains strategic elements that could confound straightforward interpretations. The key insight is that our framework learns these nuanced patterns from large-scale data, where strategic damage-taking versus skill-based vulnerability emerge as distinguishable patterns when contextualized with other gameplay features and temporal sequences, making the proxy robust despite individual behavioral variations.

While we demonstrate our framework using PUBG as our primary case study, the approach is designed to be genre-agnostic. The games used in MultiPENG (FIFA and SFV), while popular, were deliberately chosen for the availability of natural game pauses between short durations of gameplay and the explicit control over difficulty. However, they are relatively simpler games compared competitive multiplayer battle-royale games such as PUBG. We specifically selected PUBG because it represents a highly complex gaming environment that spans multiple genres (shooting, combat, scavenging, multiplayer, survival), providing an exceptionally challenging test scenario that is also commercially popular and realistic. If our framework can effectively measure engagement in PUBG's complex environment with its varied gameplay elements, it should be adaptable to less complex gaming scenarios across different genres including sports games, racing games, strategy games, and more. For instance, in sports games, skill could be measured through performance metrics like scoring efficiency or ball possession, while challenge might be quantified through opponent defensive pressure. In racing games, skill could be measured via lap times or overtaking maneuvers, while challenge might be represented by track difficulty or competitor performance. This flexibility allows our engagement measurement approach to extend beyond combat-focused games to virtually any interactive gaming experience that generates telemetry data.

When calculating skill based on player ranking, we maintained the natural composition of PUBG matches, including both human players and AI-controlled bots. This approach preserves the authentic gameplay experience, as players typically don't distinguish between human and AI opponents during combat. While bots are present in the environment, our skill and challenge validation metrics were evaluated specifically on human player data, ensuring the framework's effectiveness for measuring human engagement. Beyond training, our framework's player modeling nature effectively encodes player behavior making it applicable for AI bot detection. Furthermore, by collectively modeling player interactions, AI bots can be trained to behave according to the match settings and player composition for more realistic bots.

Dataset

Our evaluation utilizes two complementary datasets: a skill-challenge estimation dataset for model development and a survey dataset for engagement validation. Both datasets share the same underlying structure, capturing game telemetry at 10-second intervals.

For the skill-challenge estimation dataset, we implemented a systematic sampling strategy starting with five seed players (professionals, streamers, community mem-

bers), expanding to 1,684 unique players through their recent match histories. Players were categorized into five tiers based on lifetime match count using IQR-based outlier removal and quantile-based discretization, ranging from Rookies (< 374 matches, avg. 146, std. 95) to Masters ($> 6,369$ matches, avg. 10,345, std. 4,034), with Amateur (374-1,161 matches), Veteran (1,161-2,780 matches), and Elite (2,780-6,369 matches) tiers in between.

For efficient data collection, we selected four players from each tier, resulting in a balanced sample of 20 players. We monitored their battle royale matches (solo, duo, and squad modes) over a two-week period, collecting complete telemetry data for 2,673 matches. After preprocessing and phase-based segmentation, this yielded 20,030 data points, which we split into training (16,267), validation (1,866), and testing (1,897) sets.

For engagement validation, we conducted a data collection experiment involving 31 players. The experiment was approved by University of Ottawa's Office of Research Ethics and Integrity, file Number H-07-23-9439. Participants registered with their PUBG username and demographic information in our web application, then submitted post-match experiences through a structured questionnaire derived from the Game Experience Questionnaire (GEQ)[3]. For engagement measurement, participants rated their level from "Disengaged" (feeling bored, unfocused) to "Highly Engaged" (losing track of time, fully immersed), with intermediate levels of "Slightly," "Moderately," and "Fairly" engaged. Each level included descriptive examples to ensure consistent interpretation.

Our data collection workflow prioritized ecological validity - participants played PUBG matches normally, then immediately reported their skill level, perceived challenge, and engagement on 5-point Likert scales to minimize recall bias [4]. We aligned survey responses with telemetry data by matching submission timestamps with corresponding match data retrieved via the PUBG API using participants' usernames. We processed the telemetry data using the same phase-based approach, resulting in 120 labeled data points.

Following the same approach as MultiPENG, we transformed the Likert responses into binary classifications using the mean reported engagement (3.58) as the threshold and employed group-stratified cross-validation to account for person-specific reporting tendencies. These self-reported scores provided ground truth labels for framework validation. Table 4.1 summarizes the overall dataset statistics.

The Skill-Challenge dataset lacks explicit labels because it employs a self-supervised approach where complex models learn to predict skill and challenge scores directly from telemetry data of completed game sessions. This two-stage framework maximizes the utility of abundant historical gameplay data while minimizing reliance on expensive survey-collected engagement labels—only 31 engagement labels are needed to train the simpler second-stage model that maps skill-challenge scores to engagement, compared to the thousands of labels that would otherwise be required to train the complex telemetry-to-engagement model end-to-end.

Training Configuration

For PUBG implementation, we configured the framework with 6 temporal snapshots per sequence (1 minute of gameplay) and 35 player features. The categorical embeddings were dimensioned specifically for PUBG's feature cardinality: team

Table 4.1: Dataset Statistics Summary

Dataset	Matches	Datapoints	Labels
Skill-Challenge	2,673	20,030	-
Train	2,173	16,267	-
Validation	250	1,866	-
Test	250	1,897	-
Survey	31	120	31

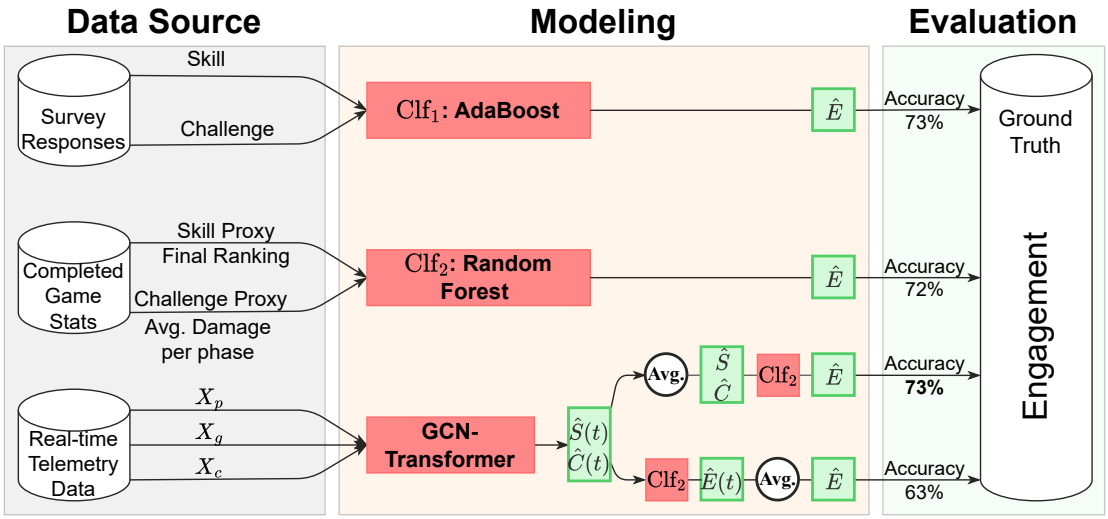


Figure 4.3: Validation framework for engagement estimation.

size (3→2), map ID (12→6), and phase index (11→4).

The GCN implementation uses two convolutional layers (64 and 16 output units) to process the player interaction graph. The Transformer encoder was configured with two layers, two attention heads, and a hidden dimension of 128. We maintained pre-layer normalization for stable training with the game's variable player counts.

Training proceeded with the AdamW optimizer (learning rate: $3e^{-4}$, weight decay: 0.01) using cosine annealing schedule. The model trained for maximum 50 epochs with early stopping (patience: 5), using batches of 128 sequences. To handle PUBG's variable player counts per match, we implemented masked loss computation for active players only.

The engagement classifier was calibrated using our PUBG survey dataset (n=31). We addressed the granularity mismatch between phase-level predictions and match-level surveys by aggregating phase estimates using final normalized ranking for skill and mean damage across phases for challenge. High engagement thresholds were determined using mean reported engagement scores (3.58±0.10). This notably high baseline engagement aligns with PUBG's status as a competitive battle royale game, where the inherent match stakes and elimination mechanics naturally foster high player investment.

4.4.2 Results & Analysis

Framework Validation

Table 4.2: Classification Performance Across Different Input Sources

Metric	Low	High	ROC_AUC	Accuracy
Stage 1: Clf_1 (Survey Responses) - AdaBoost				
Precision	0.64 ± 0.09	0.79 ± 0.04		
Recall	0.55 ± 0.08	0.88 ± 0.03	0.75 ± 0.06	0.73 ± 0.03
F1-score	0.53 ± 0.07	0.80 ± 0.03		
Stage 2: Clf_2 (Post-match Proxies) - Random Forest				
Precision	0.72 ± 0.06	0.81 ± 0.04		
Recall	0.68 ± 0.06	0.77 ± 0.05	0.76 ± 0.03	0.72 ± 0.02
F1-score	0.63 ± 0.04	0.75 ± 0.03		
Stage 3: Clf_2 (Avg. of Real-time Estimates) - Random Forest				
Precision	0.63 ± 0.04	0.87 ± 0.05		
Recall	0.88 ± 0.04	0.66 ± 0.05	**0.83 ± 0.03**	**0.73 ± 0.03**
F1-score	0.71 ± 0.03	0.73 ± 0.04		

We evaluate our engagement estimation framework through a hierarchical validation approach that progresses from theoretical foundations to practical implementation. Beginning with survey-based engagement prediction as an upper bound, we systematically validate our telemetry-based proxies before assessing the complete end-to-end framework, as illustrated in Figure 4.3.

The first stage, shown at the top of the figure, employs Clf_1, an AdaBoost classifier (empirically selected for optimal performance on survey data) that maps self-reported skill and challenge to self-reported engagement. This establishes our empirical baseline for engagement prediction from explicit player feedback. We configured AdaBoost with SAMME boosting algorithm specifically because of its effectiveness with categorical features and resilience to overfitting on small datasets like our survey responses. This stage establishes the theoretical ceiling for engagement prediction accuracy using explicit skill-challenge relationships.

The second stage, shown in the middle of the figure, evaluates our telemetry-based proxies using Clf_2, a Random Forest classifier (chosen for its superior performance on game statistics) operating on match completion statistics—final ranking for skill and average damage taken per phase for challenge. While this stage requires complete match data, it serves to validate our proxy selection methodology. This intermediate validation stage is crucial for demonstrating that our selected telemetry proxies can approach the performance of explicit player feedback, proving the viability of non-intrusive measurement. It validates that game-derived metrics can effectively replace traditional survey methods while maintaining accuracy.

Our proposed end-to-end framework, represented by the third stage shown at the bottom of the figure, combines skill-challenge measurements from the GCN-Transformer model with Clf_2. We used a hold-k-out cross-validation scheme that holds out five survey responses in each fold, maintaining complete separation between training and testing data. This real-time validation approach represents the framework's primary innovation: providing engagement estimates during game-

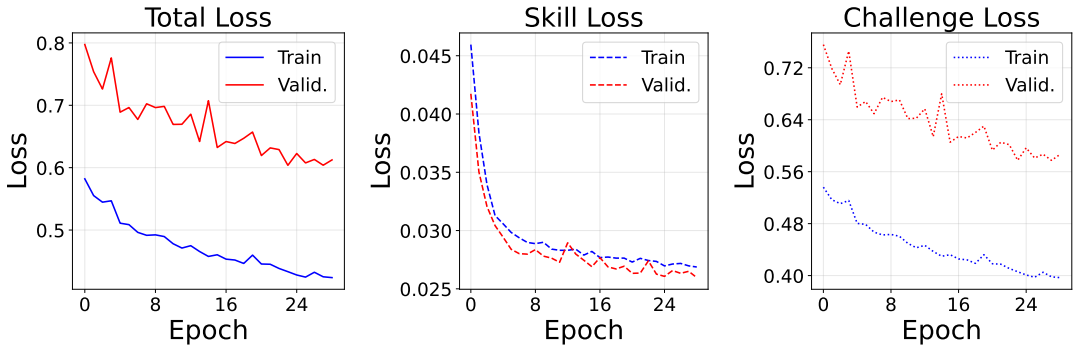


Figure 4.4: Training and validation loss curves.

play rather than retroactively. By successfully approximating match-level predictions using only partial gameplay data, our framework enables adaptive game mechanics that can respond to fluctuating engagement levels within a single match. During real-time operation, the framework processes each phase independently: measuring current skill from performance up to the current phase, measuring the challenge level for the current phase, and applying the appropriate fold-specific classifier to these phase-level measures.
Given the granularity mismatch between phase-level measures and match-level ground truths, we average phase-level skill $\hat{S}(t)$ and challenge $\hat{C}(t)$ measures before engagement classification. This approach demonstrates our framework's potential for real-world deployment across varying time scales, from moment-to-moment gameplay adaptation to session-level analytics. Game designers can leverage these multi-resolution engagement signals to optimize both immediate mechanics and broader game progression systems. We specifically avoid averaging phase-level engagement measures $\hat{E}(t)$ since Clf_2 was trained on match-level skill and challenge scores, so averaging engagement would incorrectly assume linear composition across phases, contradicting flow theory [7]. Results are shown in Table 4.2.

Training Dynamics

The learning curves (Figure 4.4) demonstrate stable convergence for both skill and challenge estimation. While the total validation loss shows some fluctuation early in training, it stabilizes around epoch 15, indicating robust model generalization. The skill component converges more quickly and shows minimal gap between training and validation performance, suggesting effective learning of ranking patterns. The challenge component exhibits a larger training-validation gap but maintains consistent improvement throughout training.

Phase-wise performance

As shown in Figure 4.5, our GCN-Transformer model's accuracy varies significantly across game phases. Skill estimation error (blue line) demonstrates a consistent improvement pattern, with RMSE decreasing steadily from 0.235 in phase 0 to 0.024 in phase 9. This trend reflects the model's increasing ability to more accurately measure player skill as more gameplay data becomes available.

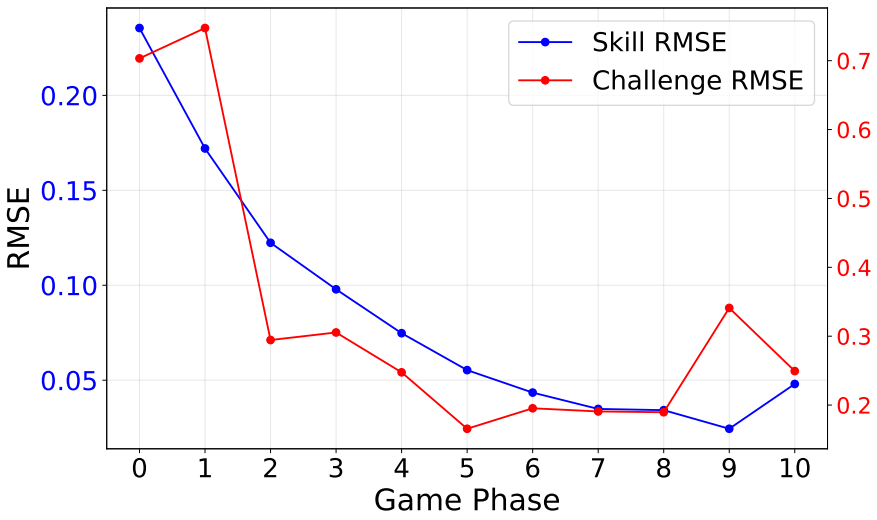


Figure 4.5: Phase-wise RMSE for skill and challenge measurement.

Challenge estimation (red line) exhibits a different pattern, with high initial error (RMSE 0.703-0.747 in phases 0-1) followed by a sharp improvement in phase 2 (RMSE 0.294). The model achieves its best challenge predictions in phase 5 (RMSE 0.166), coinciding with mid-game player confrontations as the playable area constricts. However, both skill and challenge predictions show increased error in the final phases (9-10), likely due to reduced player count and heightened end-game volatility.

These results indicate that our model's accuracy is phase-dependent, performing optimally during mid-game phases where player interactions are most structured and predictable.

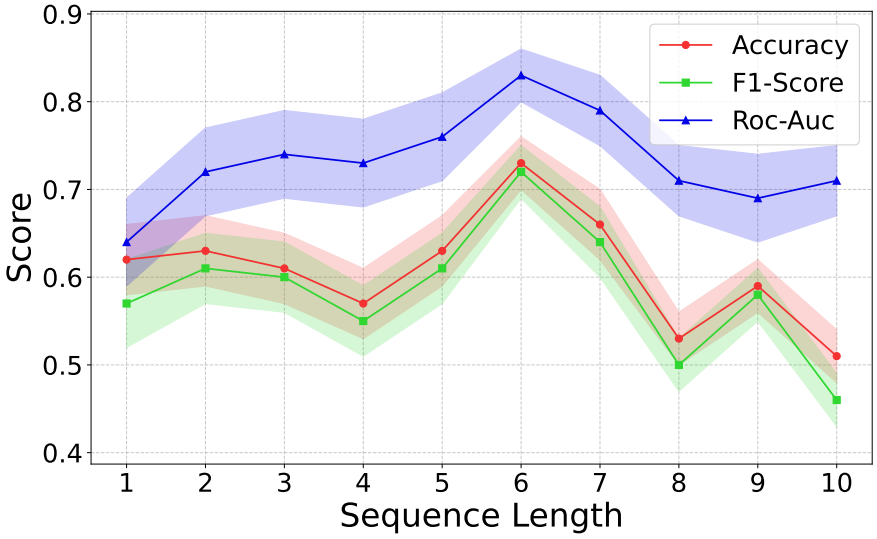


Figure 4.6: Impact of sequence length on model performance. Scores represent macro average of metric on both classes. Shaded area represents one SEM.

Sequence-Length Sensitivity

We systematically evaluated model architectures trained with different sequence lengths (1-10 timesteps at 10-second intervals) to determine the optimal temporal window for engagement measurement. As shown in Figure 4.6, all performance metrics peak at sequence length 6, with ROC-AUC reaching 0.83±0.03, achieving an accuracy of 0.73±0.03 and an F1-score of 0.71±0.03. This indicates that one minute of gameplay data is sufficient for reliable measures, which is particularly important given the short duration of game phases in PUBG.

Models trained with longer sequences not only show degraded performance but also exhibit increased variance, as evidenced by the widening standard deviation bands beyond length 7. To handle variable-length sequences in practice, our implementation employs padding when the available sequence is shorter than the target length (e.g., due to early player elimination), while longer sequences are trimmed. This approach ensures consistent input dimensionality while maintaining temporal relevance of the features.

The ability to make accurate predictions with just one minute of data enables responsive engagement measurement within the typical duration of game phases, making our framework practical for real-time applications.

As a generalization note, if the game has multiple stages and the player enters another stage, features from previous stages can be used as complementary signals depending on the similarity between the game stages, to mitigate a potential cold-start problem.

Engagement Classifier

We evaluated several classification algorithms including Random Forest, AdaBoost, and Support Vector Machines (SVM) for engagement classification. While both Random Forest and AdaBoost demonstrated competitive performance, their relative effectiveness varied based on the input features used. The Random Forest classifier emerged as the optimal choice for telemetry-based proxies, achieving an ROC-AUC of 0.83±0.03 and accuracy of 0.73±0.03.

Training on different input sources revealed distinct patterns in feature importance. With survey-based inputs, AdaBoost showed a clear bias toward challenge scores (0.622±0.01) over skill scores (0.378±0.01). In contrast, the Random Forest trained on telemetry-based proxies exhibited remarkably balanced feature importance between skill (0.502±0.01) and challenge (0.498±0.01). The minimal SEM (±0.01) across cross-validation folds indicates robust stability in this balanced relationship.

The contrast between survey-based and proxy-based feature importances reveals an interesting psychological aspect: while players may be more consciously aware of challenge levels during gameplay, our telemetry-based proxies capture a more balanced representation of the skill-challenge relationship. This finding suggests that objective gameplay metrics may better reflect the theoretical engagement model than subjective player assessments, possibly due to reporting biases or varying interpretations of skill and challenge across players.

Ablation Study

To assess the individual contribution of skill and challenge components, we conducted isolated evaluations using single-feature classifiers. The results are summarized in Table 4.3.

Table 4.3: Component-wise Classification Performance

Component	ROC-AUC	Accuracy
Skill-only	0.57 ± 0.04	0.63 ± 0.03
Challenge-only	0.69 ± 0.03	0.61 ± 0.03
Combined	**0.83 ± 0.03**	**0.73 ± 0.03**

The challenge-only classifier achieved higher ROC-AUC but lower accuracy compared to the skill-only variant, suggesting that challenge levels may be more discriminative but less reliable for binary engagement classification. However, the combined approach significantly outperformed both individual components, with improvements of 20.3% and 16.7% in ROC-AUC over skill-only and challenge-only classifiers, respectively. This substantial performance gain validates our framework's theoretical foundation in flow theory and demonstrates the synergistic relationship between skill and challenge in engagement measurement.

We also conducted extensive ablation experiments to evaluate the contribution of different feature categories, model architecture components, and assess the individual impact of skill and challenge components on the framework's performance. To validate our hybrid architecture design, we compared the full GCN-Transformer model against a Transformer-only variant that excludes the graph convolutional component. The hybrid architecture (accuracy: 0.73±0.03, ROC-AUC: 0.83±0.03) outperforms the Transformer-only model (accuracy: 0.67±0.04, ROC-AUC: 0.74±0.05), suggesting that the GCN's ability to model player interactions provides valuable information for engagement measurement. The Transformer-only model shows stronger precision for high engagement states (0.80±0.06) but suffers from reduced recall (0.63±0.05) compared to the hybrid approach.

Table 4.4 presents the impact of removing different feature categories on both skill-challenge estimation and end-to-end engagement measurement. The baseline model, utilizing all features, achieves the best performance across most metrics. Removing player features significantly degrades skill estimation (RMSE increases from 0.163 to 0.319) and end-to-end accuracy (73% to 55%). Similarly, excluding game features impairs challenge estimation (RMSE increases from 0.541 to 0.604) and reduces end-to-end accuracy to 61%. Categorical features show the least impact, with marginal changes in skill-challenge estimation and moderate degradation in end-to-end performance.

Robustness to Perceptual Variability

Figure 4.7 examines the relationship between our telemetry-based estimates and players' self-reported perceptions through Spearman's rank correlation analysis. Our skill estimates show a significant moderate correlation with self-reported skill ($\rho = 0.39$, $p = 0.03$), suggesting our ranking-based proxy effectively captures aspects of player-perceived skill. Interestingly, our challenge estimates show limited correlation with self-reported challenge ($\rho = 0.07$, $p = 0.70$). Given the

Table 4.4: Impact of Feature Removal on Model Performance

Features Removed	Skill RMSE	Challenge RMSE	Acc. (%)	ROC-AUC
none	**0.163**	**0.541**	**0.73±0.03**	**0.83±0.3**
player	0.319	0.707	0.55±0.03	0.59±0.04
game	0.343	0.604	0.61±0.03	0.56±0.03
categorical	0.172	0.535	0.57±0.04	0.72±0.04

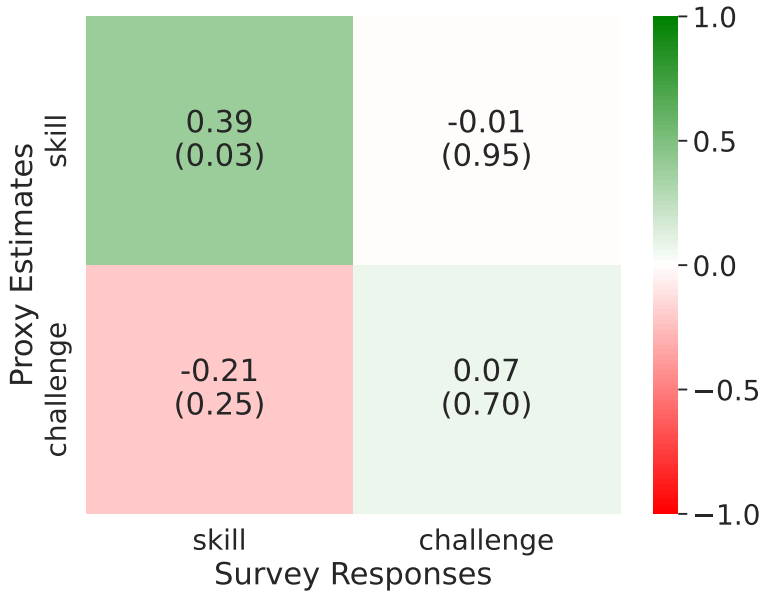


Figure 4.7: Correlation between model estimates and player survey responses

strong predictive power of our challenge proxy demonstrated in the ablation study (ROC-AUC 0.69 for challenge-only classifier), this finding suggests that effective engagement measurement may not require direct alignment with players' subjective challenge perceptions. Instead, our telemetry-based challenge metric appears to capture gameplay patterns that, while distinct from players' self-reported experiences, provide valuable signals for engagement estimation. It's important to note that the telemetry-based challenge proxy serves as a construct within the flow-based framework rather than directly capturing players' subjective perceptions of challenge. However, as demonstrated in Table 4.2, these telemetry-based proxies achieve comparable performance to survey-based measures of skill and challenge. Since the ultimate goal is measuring player engagement, what matters is that the telemetry proxies fulfill the theoretical roles of skill and challenge in flow theory, not whether they strongly correlate with players' self-reported perceptions of these constructs.

Cross-Domain Validation

To assess our framework's generalizability beyond PUBG, we considered how it might perform across different game genres. The Flow Theory-based approach demonstrated in this chapter shares its conceptual foundation with the model presented in Chapter 3, which achieved 66±2% accuracy across FIFA'23 and Street

Fighter V. As shown in that chapter, a model using only skill and challenge metrics significantly outperformed human observers and was on par with complex multimodal approaches in predicting engagement.

The findings from the MultiPENG study in Chapter 3 validate the core premise of our telemetry-based framework—that engagement can be effectively measured through the relationship between player skill and game challenge. In that study, using simple self-reported game familiarity as a skill proxy and normalized game difficulty settings as challenge metrics, the Random Forest classifier demonstrated robust performance across different game genres.

While our current implementation focuses on PUBG's complex environment, the consistent performance of Flow Theory-based models across different gaming contexts suggests our approach is genre-agnostic. The telemetry-based framework presented in this chapter extends the insights from the MultiPENG study by developing methodology to extract skill and challenge metrics directly from gameplay data rather than relying on self-reports or preset difficulty levels. This advancement enables real-time, non-intrusive engagement measurement in complex, dynamic gaming environments where traditional survey-based approaches are impractical.

For different game genres, our framework could be adapted by identifying appropriate telemetry signals that serve as skill and challenge proxies. For example, in sports games like FIFA, skill could be measured through performance metrics like scoring efficiency or ball possession, while challenge might be quantified through opponent defensive pressure. In racing games, skill could be measured via lap times or overtaking maneuvers, while challenge might be represented by track difficulty or competitor performance. The consistent performance of Flow Theory-based models across these diverse contexts, as demonstrated in the MultiPENG study, suggests that our telemetry-based approach should maintain its effectiveness when properly adapted to different game genres.

4.5 Summary and Discussion

This chapter has introduced a framework for real-time player engagement estimation that advances the state-of-the-art through its predictive capabilities and non-intrusive nature. By combining GCNs for player interaction modeling with Transformer networks for temporal processing, our framework successfully predicts skill and challenge levels before their manifestation in gameplay outcomes. This approach builds directly on the insights from the MultiPENG study in Chapter 3, which demonstrated the effectiveness of Flow Theory-based metrics in predicting engagement, and extends them into a practical, scalable implementation for complex gaming environments.

Our analysis revealed several important findings that contribute to the understanding of engagement measurement in modern video games:

First, our phase-specific analysis demonstrated that accurate engagement predictions can be made using just one minute of gameplay data, enabling responsive measurement within the typical duration of game phases. This has significant implications for the practical application of engagement measurement systems in commercial games, where timely feedback is essential for adaptive mechanics.

Second, our feature importance analysis identified player features as the most

critical for accurate estimation, followed by game state features, with categorical features having a more moderate impact. This hierarchy of feature importance provides guidance for game developers implementing streamlined versions of our framework in resource-constrained environments.

Third, we observed that objective gameplay metrics may better reflect the theoretical engagement model than subjective player assessments. This was evident in the contrast between survey-based and proxy-based feature importances, where telemetry-based proxies captured a more balanced representation of the skill-challenge relationship than players' self-reports. This finding suggests that reporting biases or varying interpretations of skill and challenge across players may limit the reliability of subjective engagement measures.

Our results empirically validate the effectiveness of Flow Theory's skill-challenge relationship in quantifying engagement within modern multiplayer games. The framework showed peak performance during structured mid-game interactions, where player relationships and behaviors are most clearly defined. This pattern aligns with Flow Theory's emphasis on clear goals and immediate feedback as key components of optimal engagement.

Importantly, the cross-domain validation discussed in relation to the MultiPENG study demonstrates that our approach extends well beyond combat games, with robust performance observed in sports games (FIFA'23) and fighting games (Street Fighter V). This suggests broad applicability across the gaming industry regardless of genre, addressing the need for generalizable engagement measurement methods identified in Chapter 2.

This chapter provides game developers and researchers with a practical, non-intrusive instrument for analyzing and monitoring player engagement using existing telemetry data. The framework's predictive capabilities enable proactive game adjustments and more efficient resource allocation in cloud gaming [13], [170]. Its scalability makes it suitable for large-scale deployment across various gaming platforms and genres, with experimental validation demonstrating robust measurement performance across different game types including sports games and fighting games.

While the framework's predictive capabilities naturally suggest applications in dynamic difficulty adjustment, this initially appears to create a problematic feedback loop—using telemetry-based challenge estimates to modify the very game elements that generate that telemetry data. However, this concern can be addressed by treating the framework as a diagnostic tool that informs targeted interventions rather than continuous automatic adjustments. For instance, if the system predicts high challenge but low engagement for a player, developers can implement discrete corrective measures such as reducing enemy strength, providing supportive items, or adjusting spawn rates. Moreover, since the framework can estimate challenge levels, developers can simulate specific adjustments beforehand to predict how modifications will affect both challenge and engagement scores, enabling data-driven optimization of difficulty interventions without disrupting ongoing gameplay.

By detecting subtle variations in engagement levels relative to an established baseline, our measurement approach offers more nuanced insights than traditional binary engagement classifications while maintaining the natural flow of gameplay [171]. This represents a significant step toward more engaging and adaptive mul-

tiplayer online games, addressing the research directions outlined in Chapter 2 for non-intrusive, real-time engagement measurement methods.
The telemetry-based framework presented in this chapter complements the multimodal approach explored in the MultiPENG study. While the MultiPENG study demonstrated the value of physiological signals for engagement detection in controlled environments, this chapter shows how engagement can be effectively measured in natural gaming environments without specialized equipment. Together, these approaches provide a comprehensive toolkit for engagement measurement across different contexts and applications, from laboratory studies to large-scale commercial deployment.

Chapter 5

Conclusion

This thesis has presented a systematic investigation of player engagement measurement in video games, progressing from theoretical foundations through experimental validation to practical implementation. The research addressed crucial gaps in engagement measurement methodology, particularly the need for comprehensive multimodal datasets with reliable engagement annotations and non-intrusive, real-time measurement approaches suitable for complex gaming environments. This concluding chapter synthesizes the key contributions, discusses limitations, and proposes directions for future research.

5.1 Summary of Contributions

This thesis has made several significant contributions to the field of player engagement measurement:

5.1.1 Comprehensive Multimodal Dataset

The MultiPENG dataset presented in Chapter 3 provides synchronized engagement data across multiple modalities (webcam footage, EEG, eye tracking, heart rate, controller inputs), enabling direct comparison of their effectiveness in identical gaming contexts. This dataset addresses a critical gap in engagement estimation research by allowing researchers to evaluate which modalities most effectively capture engagement signals. With standardized cross-validation folds and curated subsets for human evaluation, the dataset offers a valuable resource for developing and benchmarking engagement estimation methods.

5.1.2 Innovative Ground Truth Collection

The thesis introduced a methodological innovation in ground truth collection through the Experience Sampling Method implemented during natural gameplay pauses. This approach minimized both gameplay disruption and recall bias, addressing critical limitations of traditional post-hoc questionnaires and continuous self-annotation methods. By collecting engagement self-reports during organic breaks in gameplay, the method maintained ecological validity while still capturing temporal fluctuations in engagement. The validation of this approach through

correlation analysis confirmed its effectiveness for capturing distinct engagement dimensions.

5.1.3 Empirical Validation of Flow Theory

The MultiPENG study provided compelling empirical support for Flow Theory as a practical foundation for engagement estimation. The flow-based model, using only player skill and game challenge as predictors, achieved the highest overall accuracy at 67% ($\pm$0.02), on par with complex multimodal neural architectures. This finding validated the fundamental relationship between skill-challenge balance and engagement, suggesting that relatively simple, theory-driven approaches could achieve comparable or superior performance to more complex, sensor-heavy systems.

5.1.4 Real-time Measurement Framework

Chapter 4 presented a novel framework for real-time, non-intrusive engagement measurement using game telemetry data. This framework transformed standard gameplay metrics into meaningful engagement estimates, enabling practical implementation of Flow Theory-based engagement measurement without specialized equipment or gameplay interruption. The framework successfully predicted engagement using just one minute of gameplay data (6 timesteps), making it suitable for real-time applications requiring timely feedback. With an end-to-end accuracy of 73% and 0.83 ROC-AUC, the framework demonstrated superior performance to simpler alternatives while maintaining real-time processing capabilities.

5.1.5 Hybrid Technical Architecture

The telemetry-based framework introduced a novel hybrid architecture combining Graph Convolutional Networks (GCN) for modeling player interactions and spatial relationships with Transformer networks for processing temporal sequences of game states. This integrated approach outperformed single-architecture alternatives (73% vs. 67% accuracy for Transformer-only), demonstrating the value of combining spatial and temporal processing for engagement estimation. The architecture effectively captured the complex, dynamic nature of multiplayer gaming environments while maintaining computational efficiency for real-time applications.

5.1.6 Cross-Domain Validation

The thesis provided evidence for the potential cross-domain applicability of the Flow Theory-based approach. While the telemetry-based framework itself was primarily implemented and validated in PUBG, its conceptual foundation shares a common basis with the model presented in the MultiPENG study, which achieved 66$\pm$2% accuracy across FIFA'23 and Street Fighter V. The MultiPENG findings, where a model using only skill and challenge metrics significantly outperformed human observers and was comparable to complex multimodal approaches, validate the core premise of the telemetry-based framework—that engagement can

be effectively measured through the relationship between player skill and game challenge.
The key advancement in the telemetry-based framework was developing a methodology to extract skill and challenge metrics directly from gameplay data rather than relying on self-reports (for skill) or preset difficulty levels (for challenge) as was done in the MultiPENG study. This advancement enables real-time, non-intrusive engagement measurement in complex, dynamic gaming environments where traditional survey-based approaches are impractical.
The thesis discussed how the framework could be adapted to different game genres by identifying appropriate telemetry signals that serve as skill and challenge proxies. For example, in sports games like FIFA, skill might be measured through performance metrics like scoring efficiency or ball possession, while challenge could be quantified through opponent defensive pressure. In racing games, skill metrics might include lap times or overtaking maneuvers, while challenge could be represented by track difficulty or competitor performance.
The consistent performance of Flow Theory-based models across diverse contexts in the MultiPENG study suggests that the telemetry-based approach should maintain its effectiveness when properly adapted to different game genres, although direct validation across these additional genres remains an area for future work.

5.2 Practical Applications

As outlined in the introduction Section 1.7, the real-time engagement metrics provided by the framework developed in this thesis offer several practical applications for game developers and researchers:

- **Dynamic Difficulty Adjustment**: Games can automatically modify challenge levels based on detected engagement states, preventing player frustration or boredom [8], [9]
- **Targeted Content Delivery**: Developers can introduce new gameplay elements or narrative sequences precisely when engagement begins to decline [10]
- **Personalized Matchmaking**: Matchmaking systems can maintain optimal skill-challenge balances across different player segments [11], [12]
- **Intelligent Resource Allocation**: Cloud gaming environments can dynamically allocate bandwidth and processing resources to maintain quality during critical engagement periods [13]
- **Game Design Optimization**: Developers can identify which specific game elements consistently drive or diminish engagement, informing future design decisions

These applications highlight the practical value of the research beyond academic understanding, demonstrating how real-time engagement measurement can enhance player experiences and potentially improve commercial outcomes in the gaming industry.

5.3 Limitations

Despite its significant contributions, this research has several limitations that should be acknowledged:

5.3.1 Dataset Limitations

The MultiPENG dataset, while comprehensive, faced certain constraints:

- Limited demographic diversity in the participant pool, potentially affecting the generalizability of findings across different player populations
- Variation in data quality across modalities, with only 50% of samples having an average EEG quality (EQ.OVERALL) of at least 75%
- Relatively small sample size for the engagement validation component (31 players, 120 labeled data points), though this is comparable to similar studies in the field

These limitations, while common in engagement research, suggest caution in generalizing findings to all player demographics and gaming contexts.

5.3.2 Measurement Challenges

Several measurement challenges persisted throughout the research:

- Challenge estimates showed limited correlation with self-reported challenge (ρ=0.07, p=0.70) in the telemetry-based framework, suggesting that objective telemetry metrics may not fully capture subjective perceptions of difficulty
- The binary classification of engagement (high vs. low) simplifies what is likely a continuous spectrum of engagement states
- Phase-specific performance variations in the telemetry-based framework, with best accuracy during mid-game (phases 2-5), indicate context-dependent measurement effectiveness

These challenges reflect the inherent complexity of measuring subjective psychological states through objective behavioral and performance metrics.

5.3.3 Implementation Constraints

The practical implementation of the framework faced several constraints:

- Reliance on game-specific telemetry formats, requiring adaptation for different game engines and data structures
- Need for appropriate model deployment infrastructure to support the real-time capabilities demonstrated in the research

- Limited testing in live production environments with large player populations

These constraints highlight considerations for translating research prototypes into production systems.

5.4 Directions for Future Research

Building on the foundations established in this thesis, several promising directions for future research emerge:

5.4.1 Enhanced Skill and Challenge Measurement

Future research could refine skill and challenge measurement methodologies to better align with subjective player experiences:

- Developing more nuanced challenge metrics that incorporate perceived difficulty rather than just objective performance measures
- Exploring relative skill measurement that accounts for player history and improvement trajectories
- Investigating the impact of self-efficacy and confidence on the relationship between objective skill metrics and engagement
- Incorporating genre-specific skill components that capture expertise dimensions beyond traditional performance metrics

These refinements would address the limited correlation between objective challenge metrics and subjective perceptions identified in the current research.

5.4.2 Expanded Engagement Taxonomy

While Flow Theory provided an effective foundation, future work could explore a more comprehensive engagement taxonomy:

- Investigating how engagement manifests differently across various player types and demographics
- Developing modular engagement models that adapt to different game genres and mechanics
- Exploring the relationship between engagement and specific emotional states beyond the basic flow channel
- Incorporating cultural factors that influence engagement perceptions and expressions

This expanded taxonomy would provide a more nuanced understanding of how engagement operates across diverse player populations and gaming contexts.

5.4.3 Adaptive Intervention Systems

A promising application direction involves developing systems that dynamically respond to measured engagement states:

- Creating adaptive difficulty systems that maintain optimal skill-challenge balance based on real-time engagement measurements
- Developing content delivery systems that introduce new gameplay elements when engagement begins to decline
- Exploring personalized narrative pacing that adapts to individual engagement patterns
- Investigating dynamic matchmaking algorithms that optimize for engagement rather than just skill parity

These intervention systems would transform engagement measurement from an analytical tool to an active component of game design and player experience management.

5.4.4 Cross-Domain Validation and Transfer

Future research should focus on validating and extending the telemetry-based framework across different game genres:

- Directly applying the methodology to different game genres with appropriate adaptations
- Developing genre-agnostic feature extraction approaches that can work across diverse game types
- Investigating transfer learning techniques to adapt models trained in one genre to new contexts
- Establishing standardized benchmarks for engagement measurement across different game categories

This cross-domain validation would strengthen the generalizability claims and broaden the practical applications of the framework.

5.4.5 Integration with Commercial Game Analytics

To enhance practical impact, future research should focus on integration with existing game analytics systems:

- Developing standardized APIs for engagement measurement that can be implemented across different game engines
- Creating lightweight implementations suitable for mobile and browser-based games with limited telemetry capabilities

- Establishing industry benchmarks for engagement metrics across different game genres and platforms
- Investigating the relationship between engagement measurements and commercial success metrics

This integration would facilitate broader adoption of engagement measurement in commercial game development, bridging the gap between academic research and industry practice.

5.4.6 Longitudinal Engagement Patterns

While this thesis focused primarily on within-session engagement, future research could explore longer-term engagement trajectories:

- Investigating how engagement patterns evolve over multiple play sessions and throughout a game's lifecycle
- Exploring the relationship between moment-to-moment engagement and long-term retention
- Developing predictive models for player churn based on engagement pattern changes
- Examining how engagement relates to player identity formation and community participation

This longitudinal perspective would complement the real-time measurements developed in this thesis, providing a more comprehensive understanding of player engagement across different time scales.

5.5 Closing Remarks

This thesis has addressed fundamental challenges in player engagement measurement, progressing from theoretical understanding to practical implementation. By developing and validating a telemetry-based framework grounded in Flow Theory, the research offers both academic insights and practical tools for understanding and enhancing player experiences in modern gaming environments.

The findings demonstrate that effective engagement measurement need not rely on complex, multi-sensor arrays that disrupt the natural gaming experience. Instead, carefully designed metrics derived from standard game telemetry can provide accurate real-time engagement estimates, enabling new possibilities for adaptive gameplay, personalized experiences, and data-driven design decisions.

As games continue to evolve in complexity and scale, engagement measurement will play an increasingly central role in both research and development. The frameworks and methodologies presented in this thesis provide a foundation for this future work, offering a balance of theoretical rigor and practical applicability suitable for both academic researchers and industry practitioners.

The journey from player engagement theory to real-time measurement implementation underscores a broader principle in game research: that games offer unique

opportunities to study human psychology, social dynamics, and technological innovation within engaging, interactive contexts. By advancing our ability to measure engagement non-intrusively and in real-time, this thesis contributes not only to game design and development but also to our understanding of human experience in interactive digital environments.

Bibliography

[1] K. Arora, *The gaming industry: A behemoth with unprecedented global reach*, `https://www.forbes.com/councils/forbesagencycouncil/2023/11/17/the-gaming-industry-a-behemoth-with-unprecedented-global-reach`, Accessed: July 25, 2024, Nov. 2023.

[2] A. Z. Abbasi, D. H. Ting, and H. Hlavacs, "Engagement in games: Developing an instrument to measure consumer videogame engagement and its validation," *International Journal of Computer Games Technology*, vol. 2017, no. 1, p. 7 363 925, 2017.

[3] K. L. Norman, "Geq (game engagement/experience questionnaire): A review of two papers," *Interacting with computers*, vol. 25, no. 4, pp. 278–283, 2013.

[4] E. Hassan, "Recall bias can be a threat to retrospective and prospective research designs," *The Internet Journal of Epidemiology*, vol. 3, no. 2, pp. 339–412, 2006.

[5] D. Gábana Arellano, L. Tokarchuk, and H. Gunes, "Measuring affective, physiological and behavioural differences in solo, competitive and collaborative games," in *Intelligent Technologies for Interactive Entertainment*, Cham: Springer International Publishing, 2017, pp. 184–193.

[6] W. Yang, M. Rifqi, C. Marsala, and A. Pinna, "Towards better understanding of player's game experience," in *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, ser. ICMR '18, Yokohama, Japan: ACM, 2018, pp. 442–449.

[7] M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience*. New York, NY: Harper Perennial, 1990.

[8] A. Baldwin, D. Johnson, and P. A. Wyeth, "The effect of multiplayer dynamic difficulty adjustment on the player experience of video games," in *CHI'14 extended abstracts on human factors in computing systems*, Association for Computing Machinery, 2014, pp. 1489–1494.

[9] P. D. Paraschos and D. Koulouriotis, "Game difficulty adaptation and experience personalization: A literature review," *International Journal of Human–Computer Interaction*, vol. 39, pp. 1–22, 2022.

[10] M. F. Maleki and R. Zhao, "Procedural content generation in games: A survey with insights on emerging llm integration," in *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 20, 2024, pp. 167–178.

[11] M. Chen, A. N. Elmachtoub, and X. Lei, "Matchmaking strategies for maximizing player engagement in video games," *Available at SSRN 3928966*, 2021.

[12] K. Wang, H. Liu, Z. Hu, *et al.*, "Enmatch: Matchmaking for better player engagement via neural combinatorial optimization," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, 2024, pp. 9098–9106.

[13] A. A. Laghari, H. He, K. A. Memon, R. A. Laghari, I. A. Halepoto, and A. Khan, "Quality of experience (qoe) in cloud gaming models: A review," *multiagent and grid systems*, vol. 15, no. 3, pp. 289–304, 2019.

[14] A. Rashed, S. Shirmohammadi, and M. Hefeeda, "Real-time prediction of player engagement from multimodal data," *IEEE Transactions on Multimedia*, revised version in review.

[15] A. Rashed, S. Shirmohammadi, I. Amer, and M. Hefeeda, "A review of player engagement estimation in video games: Challenges and opportunities," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 21, no. 7, Jul. 2025.

[16] A. Christofferson, A. Videbaek, A. Egan, T. Rowland, and M. Madden, *Gamer survey: Young players reshape the industry*, [Accessed 22-02-2025], 2024.

[17] *Video game market report*, [Accessed 22-02-2025], 2024.

[18] S. Parvini, "Insights into the artificial intelligence ai in video games market's growth potential 2024-2033," *AP News*, 2024.

[19] K. S. Simay Karaağaç, "Gaming: Multi-platform experiences and ai integration," *Think with Google*, 2024.

[20] Institute for Operations Research and the Management Sciences, "New research analyzes video game player engagement," *ScienceDaily*, Sep. 2019.

[21] R. K. Rigney, "The evolution of sponsored streaming in the gaming industry," *Polygon*, 2024.

[22] N. Lim, *10 Types of Game Metrics and How To Use Them — sonamine.com*, `https://www.sonamine.com/blog/10-types-of-game-metrics-and-how-to-use-them`, [Accessed 22-02-2025], 2024.

[23] T. Hubka, *22 metrics all game developers should know by heart — gameanalytics.com*, `https://gameanalytics.com/blog/metrics-all-game-developers-should-know/`, [Accessed 22-02-2025], 2024.

[24] T. Hubka, *Decoding players' patterns with engagement tracing — gameanalytics.com*, `https://gameanalytics.com/blog/engagement-tracing-retention`, [Accessed 22-02-2025], 2024.

[25] K. Doherty and G. Doherty, "Engagement in hci: Conception, theory and measurement," *ACM Comput. Surv.*, vol. 51, no. 5, Nov. 2018.

[26] D. Shin, "How does immersion work in augmented reality games? a user-centric view of immersion and engagement," *Information, Communication & Society*, vol. 22, no. 9, pp. 1212–1229, 2019.

[27] M. I. Berkman and E. Akan, "Presence and immersion in virtual reality," in *Encyclopedia of Computer Graphics and Games*. Cham: Springer International Publishing, 2019, pp. 1–10.

[28] S. Huynh, S. Kim, J. Ko, R. K. Balan, and Y. Lee, "Engagemon: Multimodal engagement sensing for mobile games," *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies*, vol. 2, no. 1, pp. 1–27, 2018.

[29] E. A. Boyle, T. M. Connolly, T. Hainey, and J. M. Boyle, "Engagement in digital entertainment games: A systematic review," *Computers in human behavior*, vol. 28, no. 3, pp. 771–780, 2012.

[30] A. Voida and S. Greenberg, "Wii all play: The console game as a computational meeting place," in *Proceedings of the SIGCHI conference on human factors in computing systems*, 2009, pp. 1559–1568.

[31] N. Ducheneaut, N. Yee, E. Nickell, and R. J. Moore, "" alone together?" exploring the social dynamics of massively multiplayer online games," in *Proceedings of the SIGCHI conference on Human Factors in computing systems*, 2006, pp. 407–416.

[32] G. Ruqeyya, T. Hafeez, S. M. U. Saeed, and A. Ishwal, "Eeg-based engagement index for video game players," in *2022 International Conference on Emerging Trends in Electrical, Control, and Telecommunication Engineering (ETECTE)*, Lahore, Pakistan: IEEE, 2022, pp. 1–6.

[33] N. Bianchi-Berthouze, "Understanding the role of body movement in player engagement," *Human–Computer Interaction*, vol. 28, no. 1, pp. 40–75, 2013.

[34] D. Rae Selvig and H. Schoenau-Fog, "Non-intrusive measurement of player engagement and emotions - real-time deep neural network analysis of facial expressions during game play," in *HCI in Games*, X. Fang, Ed., Cham: Springer International Publishing, 2020, pp. 330–349.

[35] G. Guglielmo, P. M. Blom, M. Klincewicz, B. Čule, and P. Spronck, "Face in the game: Using facial action units to track expertise in competitive video game play," in *2022 IEEE Conference on Games (CoG)*, Beijing, China: IEEE, 2022, pp. 112–118.

[36] A. Winklbauer, B. Stiglbauer, M. Lankes, and M. Sporn, "Telling eyes: Linking eye-tracking indicators to affective variables," in *Proceedings of the 18th International Conference on the Foundations of Digital Games*, ser. FDG '23, Lisbon, Portugal: Association for Computing Machinery, 2023.

[37] K. Pinitas, D. Renaudie, M. Thomsen, *et al.*, "Predicting player engagement in tom clancy's the division 2: A multimodal approach via pixels and gamepad actions," in *Proceedings of the 25th International Conference on Multimodal Interaction*, ser. ICMI '23, Paris, France: ACM, 2023, pp. 488–497.

[38] W. A. IJsselsteijn, Y. A. W. de Kort, and K. Poels, *The Game Experience Questionnaire*. Eindhoven, NL: Technische Universiteit Eindhoven, 2013.

[39] D. Reguera, P. Colomer-de-Simón, I. Encinas, M. Sort, J. Wedekind, and M. Boguñá, "Quantifying human engagement into playful activities," *Scientific Reports*, vol. 10, no. 1, p. 4145, 2020.

[40] X. Chen, L. Niu, A. Veeraraghavan, and A. Sabharwal, "Faceengage: Robust estimation of gameplay engagement from user-contributed (youtube) videos," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 651–665, 2019.

[41] L. Nacke and C. Lindley, "Affective ludology, flow and immersion in a first-person shooter: Measurement of player experience," *Loading..., The Journal of the Canadian Game Studies Association*, vol. 3, no. 5, p. 21, 2009.

[42] H. L. O'Brien, I. Roll, A. Kampen, and N. Davoudi, "Rethinking (dis) engagement in human-computer interaction," *Computers in human behavior*, vol. 128, p. 107 109, 2022.

[43] S. Poeller, S. Seel, N. Baumann, and R. L. Mandryk, "Seek what you need: Affiliation and power motives drive need satisfaction, intrinsic motivation, and flow in league of legends," *Proceedings of the ACM on Human-Computer Interaction*, vol. 5, no. CHI PLAY, pp. 1–23, 2021.

[44] D.-I. D. Han, F. Melissen, and M. Haggis-Burridge, "Immersive experience framework: A delphi approach," *Behaviour & information technology*, pp. 1–17, 2023.

[45] R. M. Ryan and E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," *Contemporary educational psychology*, vol. 25, no. 1, pp. 54–67, 2000.

[46] E. N. Wiebe, A. Lamb, M. Hardy, and D. Sharek, "Measuring engagement in video game-based environments: Investigation of the user engagement scale," *Computers in Human Behavior*, vol. 32, pp. 123–132, 2014.

[47] D. Weibel and B. Wissmath, "Immersion in computer games: The role of spatial presence and flow," *International Journal of Computer Games Technology*, vol. 2011, Jan. 2011.

[48] K. Procci, "The subjective gameplay experience: An examination of the revised game engagement model," Ph.D. dissertation, University of Central Florida, 2015.

[49] T. Terkildsen and G. Makransky, "Measuring presence in video games: An investigation of the potential use of physiological measures as indicators of presence," *International Journal of Human-Computer Studies*, vol. 126, pp. 64–80, 2019.

[50] J. C. Read, S. MacFarlane, and C. Casey, "Endurability, engagement and expectations: Measuring children's fun," in *Interaction design and children*, vol. 2, Eindhoven, Netherlands: Shaker Publishing, 2002, pp. 1–23.

[51] H. L. O'Brien and E. G. Toms, "What is user engagement? a conceptual framework for defining user engagement with technology," *Journal of the American society for Information Science and Technology*, vol. 59, no. 6, pp. 938–955, 2008.

[52] G. N. Yannakakis, P. Spronck, D. Loiacono, and E. André, "Player Modeling," in *Artificial and Computational Intelligence in Games*, ser. Dagstuhl Follow-Ups, S. M. Lucas, M. Mateas, M. Preuss, P. Spronck, and J. Togelius, Eds., vol. 6, Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2013, pp. 45–59.

[53] S. Reis, L. P. Reis, and N. Lau, "Player engagement enhancement with video games," in *New Knowledge in Information Systems and Technologies*, Á. Rocha, H. Adeli, L. P. Reis, and S. Costanzo, Eds., Cham: Springer International Publishing, 2019, pp. 263–272.

[54] G. Yannakakis and J. Togelius, "Experience-driven procedural content generation," *IEEE Transactions on Affective Computing*, vol. 2, no. 3, pp. 147–161, 2011.

[55] L. A. Feldman, "Valence focus and arousal focus: Individual differences in the structure of affective experience.," *Journal of personality and social psychology*, vol. 69, no. 1, p. 153, 1995.

[56] K. Isbister and N. Schaffer, *Game usability: Advancing the player experience*. Boston: Morgan Kaufmann, 2008.

[57] A. Drachen, A. Canossa, and G. N. Yannakakis, "Player modeling using self-organization in tomb raider: Underworld," in *2009 IEEE Symposium on Computational Intelligence and Games*, Milan, Italy: IEEE, 2009, pp. 1–8.

[58] S. Tekofsky, P. Spronck, A. Plaat, H. van den Herik, and J. Broersen, "Psyops: Personality assessment through gaming behavior," English, in *Proceedings of the BNAIC conference 2013*, NL: Technische Universiteit, 2013, pp. 354–355.

[59] M. Szwoch and W. Szwoch, "Emotion recognition for affect aware video games," in *Image Processing & Communications Challenges 6*, R. S. Choraś, Ed., Cham: Springer International Publishing, 2015, pp. 227–236.

[60] A. Miguel-Cruz, A. M. R. Rincon, C. Daum, *et al.*, "Predicting engagement in older adults with and without dementia while playing mobile games," *IEEE Instrumentation & Measurement Magazine*, vol. 24, no. 6, pp. 29–36, 2021.

[61] A. Mehrabian, *Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies*. Oelgeschlager, Gunn & Hain, 1980.

[62] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," *Presence*, vol. 7, no. 3, pp. 225–240, 1998.

[63] I. Kniestedt, I. Lefter, S. Lukosch, and F. M. Brazier, "Re-framing engagement for applied games: A conceptual framework," *Entertainment Computing*, vol. 41, p. 100 475, 2022.

[64] E. Brown and P. Cairns, "A grounded investigation of game immersion," in *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '04, Vienna, Austria: ACM, 2004, pp. 1297–1300.

[65] L. Ermi and F. Mäyrä, "Fundamental components of the gameplay experience: Analyzing immersion," *Worlds in play: International perspectives on digital games research*, vol. 21, p. 37, 2007.

[66] C. Jennett, A. L. Cox, P. Cairns, *et al.*, "Measuring and defining the experience of immersion in games," *International journal of human-computer studies*, vol. 66, no. 9, pp. 641–661, 2008.

[67] A. McMahan, "Immersion, engagement, and presence: A method for analyzing 3-d video games," in *The video game theory reader*, M. J. Wolf and B. Perron, Eds., Routledge, 2013, pp. 67–86.

[68] L. Caroux, "Presence in video games: A systematic review and meta-analysis of the effects of game design choices," *Applied Ergonomics*, vol. 107, p. 103 936, 2023.

[69] C. C. Bracken and P. Skalski, "Presence and video games: The impact of image quality and skill level," in *Proceedings of the Ninth Annual International Workshop on Presence*, OH, USA: Cleveland State University, Aug. 2006, pp. 28–29.

[70] D. Wilcox-Netepczuk, "Immersion and realism in video games - the confused moniker of video game engrossment," in *Proceedings of CGAMES'2013 USA*, KY, USA: IEEE, 2013, pp. 92–95.

[71] J. L. Zaichkowsky, "Measuring the involvement construct," *Journal of consumer research*, vol. 12, no. 3, pp. 341–352, 1985.

[72] P. Sweetser and P. Wyeth, "Gameflow: A model for evaluating player enjoyment in games," *Comput. Entertain.*, vol. 3, no. 3, p. 3, Jul. 2005.

[73] B. Cowley, D. Charles, M. Black, and R. Hickey, "Toward an understanding of flow in video games," *Computers in Entertainment (CIE)*, vol. 6, no. 2, pp. 1–27, 2008.

[74] G. Chanel, C. Rebetez, M. Bétrancourt, and T. Pun, "Boredom, engagement and anxiety as indicators for adaptation to difficulty in games," in *Proceedings of the 12th International Conference on Entertainment and Media in the Ubiquitous Era*, ser. MindTrek '08, Tampere, Finland: ACM, 2008, pp. 13–17.

[75] J. Chen, "Flow in games (and everything else)," *Commun. ACM*, vol. 50, no. 4, pp. 31–34, Apr. 2007.

[76] G. B. Moneta, "On the measurement and conceptualization of flow," *Advances in flow research*, pp. 23–50, 2012.

[77] M. Bassi and A. Delle Fave, "Flow in the context of daily experience fluctuation," in *Flow Experience: Empirical Research and Applications*. Cham: Springer International Publishing, 2016, pp. 181–196.

[78] W. Dember and L. Penwell, "Happiness, depression, and the pollyanna principle," *Bulletin of the Psychonomic Society*, vol. 15, no. 5, pp. 321–323, 1980.

[79] A. Rashed, S. Shirmohammadi, and M. Hefeeda, "Real-time player engagement measurement using non-intrusive game telemetry," *IEEE Open Journal of Instrumentation and Measurement*, vol. 4, pp. 1–16, 2025.

[80] S. Fairclough, "Psychophysiological inference and physiological computer games," *BRAINPLAY 07 Brain-computer Interfaces and Games Workshop at Advances in Computer Entertainment (AcE)*, vol. 7, p. 6, Jan. 2007.

[81] D. Weiler et al., "Wearable heart rate monitor technology accuracy in research: A comparative study between ppg and ecg technology," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, pp. 1292–1296, Sep. 2017.

[82] R. L. Mandryk and K. M. Inkpen, "Physiological indicators for the evaluation of co-located collaborative play," in *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW '04, IL, USA: ACM, 2004, pp. 102–111.

[83] M. Bennett, L. Čironis, A. Sousa, *et al.*, "Continuous monitoring of hrv in esports players," *International Journal of Esports*, vol. 1, no. 1, 2022.

[84] J.-Y. Kim, H.-S. Kim, D.-J. Kim, S.-K. Im, and M.-S. Kim, "Identification of video game addiction using heart-rate variability parameters," *Sensors*, vol. 21, no. 14, p. 4683, 2021.

[85] M. R. Welsh, E. Mosley, S. Laborde, *et al.*, "The use of heart rate variability in esports: A systematic review," *Psychology of Sport and Exercise*, vol. 69, p. 102 495, 2023.

[86] S. Tognetti, M. Garbarino, A. Bonarini, and M. Matteucci, "Modeling enjoyment preference from physiological responses in a car racing game," in *Proceedings of the 2010 IEEE Conference on Computational Intelligence and Games*, Copenhagen, Denmark: IEEE, 2010, pp. 321–328.

[87] A. Fortin-Côté, C. Chamberland, M. Parent, *et al.*, "Predicting video game players' fun from physiological and behavioural data," in *Advances in Information and Communication Networks*, Cham: Springer International Publishing, 2019, pp. 479–495.

[88] M. Granato, D. Gadia, D. Maggiorini, and L. A. Ripamonti, "Emotions detection through the analysis of physiological information during video games fruition," in *Games and Learning Alliance*, Cham: Springer International Publishing, 2017, pp. 197–207.

[89] C. Politowski et al., "Improving engagement assessment in gameplay testing sessions using iot sensors," in *Proceedings of the IEEE/ACM 42nd International Conference on Software Engineering Workshops*, ser. ICSEW'20, Seoul, Republic of Korea: ACM, 2020, pp. 655–659.

[90] E. J. Pretty, R. Guarese, C. A. Dziego, H. M. Fayek, and F. Zambetta, "Multimodal measurement of cognitive load in a video game context: A comparative study between subjective and objective metrics," *IEEE Transactions on Games*, pp. 1–14, 2024.

[91] R. L. Hazlett, "Measuring emotional valence during interactive experiences: Boys at video game play," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06, Montréal, Québec, Canada: ACM, 2006, pp. 1023–1026.

[92] T. Bjørner, "Using eeg data as dynamic difficulty adjustment in a serious game about the plastic pollution in the oceans," in *Proceedings of the 2023 ACM Conference on Information Technology for Social Good*, ser. GoodIT '23, Lisbon, Portugal: ACM, 2023, pp. 6–15.

[93] R. Berta, F. Bellotti, A. De Gloria, D. Pranantha, and C. Schatten, "Electroencephalogram and physiological signal analysis for assessing flow in games," *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 5, no. 2, pp. 164–175, 2013.

[94] T. D. Parsons, T. McMahan, and I. Parberry, "Classification of video game player experience using consumer-grade electroencephalography," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 3–15, 2020.

[95] A. Kamzanova, G. Matthews, A. Kustubayeva, and S. Jakupov, "Eeg indices to time-on-task effects and to a workload manipulation (cueing)," *International Journal of Psychological and Behavioral Sciences*, vol. 5, no. 8, pp. 928–931, 2011.

[96] A. Gevins, M. E. Smith, L. McEvoy, and D. Yu, "High-resolution eeg mapping of cortical activation related to working memory: Effects of task difficulty, type of processing, and practice.," *Cerebral cortex (New York, NY: 1991)*, vol. 7, no. 4, pp. 374–385, 1997.

[97] J. A. Miller, U. Narayan, M. Hantsbarger, S. Cooper, and M. S. El-Nasr, "Expertise and engagement: Re-designing citizen science games with players' minds in mind," in *Proceedings of the 14th International Conference on the Foundations of Digital Games*, ser. FDG '19, CA, USA: ACM, 2019.

[98] T. Hafeez, S. M. Umar Saeed, A. Arsalan, S. M. Anwar, M. U. Ashraf, and K. Alsubhi, "Eeg in game user analysis: A framework for expertise classification during gameplay," *Plos one*, vol. 16, no. 6, e0246913, 2021.

[99] I. Rejer and M. Twardochleb, "Gamers' involvement detection from eeg data with cgaam–a method for feature selection for clustering," *Expert Systems with Applications*, vol. 101, pp. 196–204, 2018.

[100] Emotiv, *Performance metrics*, Accessed: 22-Feb-2025, Emotiv, 2023.

[101] T. Killedar et al., "Fuzzy logic for video game engagement analysis using facial emotion recognition," in *2021 8th International Conference on Signal Processing and Integrated Networks (SPIN)*, 26-27 August 2021: IEEE, 2021, pp. 481–485.

[102] G. Schiavo, A. Cappelletti, and M. Zancanaro, "Engagement recognition using easily detectable behavioral cues," *Intelligenza Artificiale*, vol. 8, pp. 197–210, 2014.

[103] P. Mavromoustakos-Blom, M. Kosa, S. Bakkes, and P. Spronck, "Correlating facial expressions and subjective player experiences in competitive hearthstone," in *Proceedings of the 16th International Conference on the Foundations of Digital Games*, ser. FDG '21, Montreal, QC, Canada: ACM, 2021.

[104] A. Toisoul, J. Kossaifi, A. Bulat, G. Tzimiropoulos, and M. Pantic, "Estimation of continuous valence and arousal levels from faces in naturalistic conditions," *Nature Machine Intelligence*, vol. 3, no. 1, pp. 42–50, 2021.

[105] J. Wiggins, M. Kulkarni, W. Min, *et al.*, "Affect-based early prediction of player mental demand and engagement for educational games," *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 14, no. 1, pp. 243–249, Sep. 2018.

[106] F. Iqbal, "Understanding user interaction in a video game by using eye tracking and facial expressions analysis," M.S. thesis, University of Tampere, School of Information Sciences, 2015.

[107] S. Wang, X. Xiong, Y. Xu, *et al.*, "Face-tracking as an augmented input in video games: Enhancing presence, role-playing and control," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '06, Montréal, Québec, Canada: ACM, 2006, pp. 1097–1106.

[108] J. Juvrud, G. Ansgariusson, P. Selleby, and M. Johansson, "Game or watch: The effect of interactivity on arousal and engagement in video game media," *IEEE Transactions on Games*, vol. 14, no. 2, pp. 308–317, 2021.

[109] W. Lu, H. He, A. Urban, and J. Griffin, "What the eyes can tell: Analyzing visual attention with an educational video game," in *ACM Symposium on Eye Tracking Research and Applications*, ser. ETRA '21 Short Papers, Virtual Event, Germany: ACM, 2021.

[110] M. Burch and K. Kurzhals, "Visual analysis of eye movements during game play," in *ACM Symposium on Eye Tracking Research and Applications*, 2020, pp. 1–5.

[111] M. Ninaus, K. Kiili, G. Wood, K. Moeller, and S. E. Kober, "To add or not to add game elements? exploring the effects of different cognitive task designs using eye tracking," *IEEE Transactions on Learning Technologies*, vol. 13, no. 4, pp. 847–860, 2020.

[112] E. Wood, T. Baltruaitis, X. Zhang, Y. Sugano, P. Robinson, and A. Bulling, "Rendering of eyes for eye-shape registration and gaze estimation," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile: IEEE, 2015, pp. 3756–3764.

[113] K. Makantasis, A. Liapis, and G. N. Yannakakis, "From pixels to affect: A study on games and player experience," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, Cambridge, UK: IEEE, 2019, pp. 1–7.

[114] K. Makantasis, A. Liapis, and G. N. Yannakakis, "The pixels and sounds of emotion: General-purpose representations of arousal in games," *IEEE Transactions on Affective Computing*, vol. 14, no. 1, pp. 680–693, 2021.

[115] D. Melhart, D. Gravina, and G. N. Yannakakis, "Moment-to-moment engagement prediction through the eyes of the observer: Pubg streaming on twitch," in *Proceedings of the 15th International Conference on the Foundations of Digital Games*, ser. FDG '20, Bugibba, Malta: Association for Computing Machinery, 2020.

[116] D. Melhart, A. Azadvar, A. Canossa, A. Liapis, and G. N. Yannakakis, "Your gameplay says it all: Modelling motivation in tom clancy's the division," in *2019 IEEE Conference on Games (CoG)*, London, UK: IEEE, 2019, pp. 1–8.

[117] H. Duan, Y. Huang, Y. Zhao, Z. Huang, and W. Cai, "User-generated content and editors in video games: Survey and vision," in *2022 IEEE conference on games (CoG)*, IEEE, 2022, pp. 536–543.

[118] J. Peña and J. T. Hancock, "An analysis of socioemotional and task communication in online multiplayer video games," *Communication research*, vol. 33, no. 1, pp. 92–109, 2006.

[119] C. Bailey, E. Pearson, S. Gkatzidou, and S. Green, "Using video games to develop social, collaborative and communication skills," in *EdMedia+ Innovate Learning*, Association for the Advancement of Computing in Education (AACE), 2006, pp. 1154–1161.

[120] *Player Engagement Metrics — larksuite.com*, `https://www.larksuite.com/en_us/topics/gaming-glossary/player-engagement-metrics`, [Accessed 22-02-2025], 2024.

[121] M. Mohammadpoor Faskhodi, M. Fernández-Chimeno, and M. A. García-González, "Arousal detection by using ultra-short-term heart rate variability (hrv) analysis," *Frontiers in Medical Engineering*, vol. 1, p. 1 209 252, 2023.

[122] J. H. Brockmyer, C. M. Fox, K. A. Curtiss, E. McBroom, K. M. Burkhart, and J. N. Pidruzny, "The development of the game engagement questionnaire: A measure of engagement in video game-playing," *Journal of experimental social psychology*, vol. 45, no. 4, pp. 624–634, 2009.

[123] A. Denisova, A. I. Nordin, and P. Cairns, "The convergence of player experience questionnaires," in *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*, TX, USA: ACM, 2016, pp. 33–37.

[124] A. I. Nordin, A. Denisova, and P. Cairns, "Too many questionnaires: Measuring player experience whilst playing digital games," in *The Seventh York Doctoral Symposium on Computer Science and Electronics*, York, UK: University of York, Sep. 2014, p. 6.

[125] R. M. Ryan, C. S. Rigby, and A. Przybylski, "The motivational pull of video games: A self-determination theory approach," *Motivation and emotion*, vol. 30, pp. 344–360, 2006.

[126] Y. A. De Kort, W. A. IJsselsteijn, and K. Poels, "Digital games as social presence technology: Development of the social presence in gaming questionnaire (spgq)," *Proceedings of PRESENCE*, vol. 195203, pp. 1–9, 2007.

[127] P. Lopes, G. N. Yannakakis, and A. Liapis, "Ranktrace: Relative and unbounded affect annotation," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, TX, USA: IEEE, 2017, pp. 158–163.

[128] D. Melhart, A. Liapis, and G. N. Yannakakis, "The arousal video game annotation (again) dataset," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 2171–2184, 2022.

[129] F. Ringeval, A. Sonderegger, J. Sauer, and D. Lalanne, "Introducing the recola multimodal corpus of remote collaborative and affective interactions," in *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, Shanghai, China: IEEE, 2013, pp. 1–8.

[130] J. Kossaifi, R. Walecki, Y. Panagakis, *et al.*, "Sewa db: A rich database for audio-visual emotion and sentiment research in the wild," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 3, pp. 1022–1040, 2019.

[131] P. Mavromoustakos-Blom, D. Melhárt, A. Liapis, G. N. Yannakakis, S. Bakkes, and P. Spronck, "Multiplayer tension in the wild: A hearthstone case," in *Proceedings of the 18th International Conference on the Foundations of Digital Games*, ser. FDG '23, Lisbon, Portugal: ACM, 2023.

[132] B. Hoffman and L. Nadelson, "Motivational engagement and video gaming: A mixed methods study," *Educational Technology Research and Development*, vol. 58, pp. 245–270, 2010.

[133] H. Schoenau-Fog and T. Bjørner, "'sure, i would like to continue': A method for mapping the experience of engagement in video games," *Bulletin of Science, Technology & Society*, vol. 32, no. 5, pp. 405–412, 2012.

[134] B. Bontchev and D. Vassileva, "Assessing engagement in an emotionally-adaptive applied game," in *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, ser. TEEM '16, Salamanca, Spain: Association for Computing Machinery, 2016, pp. 747–754.

[135] H. Schoenau-Fog, "The player engagement process–an exploration of continuation desire in digital games," in *Proceedings of DiGRA 2011 Conference: Think Design Play*, Hilversum, The NL: Digital Games Research Association (DiGRA), 2011, p. 18.

[136] M. Barthet, M. Kaselimi, K. Pinitas, K. Makantasis, A. Liapis, and G. N. Yannakakis, "Gamevibe: A multimodal affective game corpus," *Scientific Data*, vol. 11, no. 1, p. 1306, 2024.

[137] K. Xie, V. W. Vongkulluksn, B. C. Heddy, and Z. Jiang, "Experience sampling methodology and technology: An approach for examining situational, longitudinal, and multi-dimensional characteristics of engagement," *Educational technology research and development*, vol. 72, no. 5, pp. 2585–2615, 2024.

[138] R. Sawyer, J. Rowe, R. Azevedo, and J. Lester, "Modeling player engagement with bayesian hierarchical models," *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, vol. 14, no. 1, pp. 257–263, Sep. 2018.

[139] S. Pan, G. J. Xu, K. Guo, S. H. Park, and H. Ding, "Video-based engagement estimation of game streamers: An interpretable multimodal neural network approach," *IEEE Transactions on Games*, pp. 1–12, 2023.

[140] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, vol. 97, CA, USA: PMLR, Jun. 2019, pp. 6105–6114.

[141] A. Rashed, S. Shirmohammadi, and M. Hefeeda, "Descriptor: Multimodal dataset for player engagement analysis in video games (multipeng)," *IEEE Data Descriptions*, vol. 2, pp. 17–25, 2025.

[142] P. Buono, B. De Carolis, F. D'Errico, N. Macchiarulo, and G. Palestra, "Assessing student engagement from facial behavior in on-line learning," *Multimedia Tools and Applications*, vol. 82, no. 9, pp. 12 859–12 877, 2023.

[143] C. T. Tan, S. Bakkes, and Y. Pisan, "Inferring player experiences using facial expressions analysis," in *Proceedings of the 2014 Conference on Interactive Entertainment*, 2014, pp. 1–8.

[144] T. Baltrušaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "Openface 2.0: Facial behavior analysis toolkit," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, IEEE, 2018, pp. 59–66.

[145] A. R. Clarke, R. J. Barry, R. McCarthy, M. Selikowitz, and S. J. Johnstone, "Effects of stimulant medications on the eeg of girls with attention-deficit/hyperactivity disorder," *Clinical Neurophysiology*, vol. 118, no. 12, pp. 2700–2708, 2007.

[146] D. Rajput, W.-J. Wang, and C.-C. Chen, "Evaluation of a decided sample size in machine learning applications," *BMC Bioinformatics*, vol. 24, no. 1, p. 48, Feb. 2023.

[147] H.-C. Chiang, Y.-H. Wu, G.-H. Li, S. Shirmohammadi, and C.-H. Hsu, "Palm: Personalized active learning for mmwave-based activity recognition," *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–14, 2025.

[148] N. Kokhlikyan, V. Miglani, M. Martin, *et al.*, *Captum: A unified and generic model interpretability library for pytorch*, 2020.

[149] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proceedings of the 34th International Conference on Machine Learning - Volume 70*, ser. ICML'17, Sydney, NSW, Australia: JMLR.org, 2017, pp. 3319–3328.

[150] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: A review," *Journal of neural engineering*, vol. 16, no. 3, p. 031 001, 2019.

[151] L. A. Gil-Aciron, "The gamer psychology: A psychological perspective on game design and gamification," *Interactive Learning Environments*, vol. 32, no. 1, pp. 183–207, 2024.

[152] Z. Yu, M. Gao, and L. Wang, "The effect of educational games on learning outcomes, student motivation, engagement and satisfaction," *Journal of Educational Computing Research*, vol. 59, no. 3, pp. 522–546, 2021.

[153] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting, and C. Bauckhage, "Predicting player churn in the wild," in *2014 IEEE Conference on Computational Intelligence and Games*, Dortmund, Germany: IEEE, 2014, pp. 1–8.

[154] V. Bonometti, C. Ringer, M. Hall, A. Wade, and A. Drachen, "Modelling early user-game interactions for joint estimation of survival time and churn probability," in *2019 IEEE Conference on Games (CoG)*, London, UK: IEEE, 2019, pp. 1–8.

[155] T. H. Laine and R. S. N. Lindberg, "Designing engaging games for education: A systematic literature review on game motivators and design principles," *IEEE Transactions on Learning Technologies*, vol. 13, no. 4, pp. 804–821, 2020.

[156] X. Zhong and J. Xu, "Game updates enhance players' engagement: A case of dota2," in *Proceedings of the 4th International Conference on Information Management and Management Science*, 2021, pp. 117–123.

[157] G. Hookham and K. Nesbitt, "A systematic review of the definition and measurement of engagement in serious games," in *Proceedings of the Australasian Computer Science Week Multiconference*, ser. ACSW 2019, Sydney, NSW, Australia: Association for Computing Machinery, 2019.

[158] M.-V. Aponte, G. Levieux, and S. Natkin, "Scaling the level of difficulty in single player video games," in *Entertainment Computing – ICEC 2009*, S. Natkin and J. Dupire, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 24–35.

[159] D. Wheat, M. Masek, C. P. Lam, and P. Hingston, "Modeling perceived difficulty in game levels," in *Proceedings of the Australasian Computer Science Week Multiconference*, ser. ACSW '16, Canberra, Australia: Association for Computing Machinery, 2016.

[160] N. M. Diah, A. P. Sutiono, L. Zuo, *et al.*, "Quantifying engagement of video games: Pac-man and dota (defense of the ancients)," in *17th International Conference on Mathematical and Computational Methods in Science and Engineering (MACMESE15)*, Kuala Lumpur: WSEAS, 2015, pp. 49–55.

[161] A. Apicella, P. Arpaia, G. Mastrati, N. Moccaldi, and R. Prevete, "Preliminary validation of a measurement system for emotion recognition," in *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2020, pp. 1–6.

[162] A. Eddin Alchalabi, M. Elsharnouby, S. Shirmohammadi, and A. Nour Eddin, "Feasibility of detecting adhd patients' attention levels by classifying their eeg signals," in *2017 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2017, pp. 314–319.

[163] M. Catelani, L. Ciani, and C. Risaliti, "Risk assessment in the use of medical devices: A proposal to evaluate the impact of the human factor," in *2014 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2014, pp. 1–6.

[164] A. Lochbihler, B. Wallace, K. V. Benthem, *et al.*, "Assessing driver engagement through machine learning classification of physiological measures," in *2023 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2023, pp. 1–6.

[165] F. Dobrian, V. Sekar, A. Awan, *et al.*, "Understanding the impact of video quality on user engagement," *ACM SIGCOMM computer communication review*, vol. 41, no. 4, pp. 362–373, 2011.

[166] P. Xenopoulos and C. Silva, "Graph neural networks to predict sports outcomes," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021, pp. 1757–1763.

[167] M. Stöckl, T. Seidl, D. Marley, and P. Power, "Making offensive play predictable -using a graph convolutional network to understand defensive performance in soccer," Apr. 2021.

[168] I. Simpson, R. J. Beal, D. Locke, and T. J. Norman, "Seq2event: Learning the language of soccer using transformer-based match event prediction," in *Proceedings of the 28th ACM sigkdd conference on knowledge discovery and data mining*, 2022, pp. 3898–3908.

[169] T. Baltrusaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 2, pp. 423–443, Feb. 2019.

[170] I. Slivar, L. Skorin-Kapov, and M. Suznjevic, "Qoe-aware resource allocation for multiple cloud gaming users sharing a bottleneck link," in *2019 22nd conference on innovation in clouds, internet and networks and workshops (ICIN)*, IEEE, 2019, pp. 118–123.

[171] K. Xiaohan, M. N. A. Khalid, and H. Iida, "Player satisfaction model and its implication to cultural change," *IEEE Access*, vol. 8, pp. 184 375–184 382, 2020.

[172] P. Ekman and W. V. Friesen, "Facial action coding system," *Environmental Psychology & Nonverbal Behavior*, 1978.

Appendix A

MultiPENG Dataset Features

This appendix documents the feature set used in the MultiPENG dataset across three sensing modalities: electroencephalography (EEG), eye tracking, and facial expression and head movement analysis.

A.1 Electroencephalography (EEG) Features

The EEG features follow the naming convention `POW.[Channel].[Frequency Band]`. The dataset includes 70 EEG features (14 channels $\times$ 5 frequency bands). The raw EEG voltage data was collected at 128 Hz, while the band power features were sampled at 8 Hz.

EEG Channel Locations:

- **Frontal Region:** AF3, AF4 (Anterior Frontal); F3, F4 (Frontal); F7, F8 (Lateral Frontal); FC5, FC6 (Fronto-Central)
- **Temporal Region:** T7, T8
- **Parietal Region:** P7, P8
- **Occipital Region:** O1, O2

EEG Frequency Bands:

- **Theta:** 4-8 Hz
- **Alpha:** 8-12 Hz
- **Low Beta (BetaL):** 12-16 Hz
- **High Beta (BetaH):** 16-25 Hz
- **Gamma:** 25-45 Hz

Feature Generation Pattern: The EEG features are systematically named by combining each channel with each frequency band. For example:

- `POW.AF3.Theta`: Power in the Theta band at the AF3 electrode
- `POW.F4.Alpha`: Power in the Alpha band at the F4 electrode
- `POW.O2.Gamma`: Power in the Gamma band at the O2 electrode

A.2 Eye Tracking Features

Eye tracking data was sampled at 60 Hz. The following features were extracted:
Gaze Position Features:

- `FPOGX`: X-coordinate of fixation point of gaze (0-1, percentage of screen width)
- `FPOGY`: Y-coordinate of fixation point of gaze (0-1, percentage of screen height)
- `FPOGD`: Duration of the current fixation (seconds)

Eye Movement Features:

- `SACCADE_MAG`: Magnitude of saccade, calculated as distance between current and previous fixation
- `SACCADE_DIR`: Direction of saccade, calculated as angle of vector from horizontal (radians)

Pupil Features:

- `LPD`: Left pupil diameter (pixels)
- `RPD`: Right pupil diameter (pixels)

Blinking Features:

- `is_blinking`: Binary indicator (0/1) of current blinking
- `BKDUR`: Duration of the preceding blink (seconds)
- `BKPMIN`: Number of blinks in the previous 60-second period

A.3 Facial Expression and Head Movement Features

Facial expression and head movement data was extracted using OpenFace at 30 frames per second (30 Hz).
Head Pose Features:

- Position: `pose_Tx`, `pose_Ty`, `pose_Tz` - head location relative to camera (mm)
- Rotation: `pose_Rx` (pitch), `pose_Ry` (yaw), `pose_Rz` (roll) - in radians

Motion Derivative Features:
First derivatives (velocity) and second derivatives (acceleration) were calculated for all head pose features:

- Position velocity: `velocity_pose_Tx`, `velocity_pose_Ty`, `velocity_pose_Tz`

- Position acceleration: `acceleration_pose_Tx`, `acceleration_pose_Ty`, `acceleration_pose_Tz`
- Rotation velocity: `velocity_pose_Rx`, `velocity_pose_Ry`, `velocity_pose_Rz`
- Rotation acceleration: `acceleration_pose_Rx`, `acceleration_pose_Ry`, `acceleration_pose_Rz`

The dataset includes intensity measurements for 17 Facial Action Units based on the Facial Action Coding System (FACS) [172]. Each AU is represented with the suffix _r indicating intensity (range 0-5).

Table A.1: Facial Action Units

AU	**FACS Name**	**Description**
AU01	Inner Brow Raiser	Raises the inner portion of the eyebrows
AU02	Outer Brow Raiser	Raises the outer portion of the eyebrows
AU04	Brow Lowerer	Lowers and draws brows together
AU05	Upper Lid Raiser	Widens the eyes
AU06	Cheek Raiser	Raises the cheeks
AU07	Lid Tightener	Tightens the eyelids
AU09	Nose Wrinkler	Wrinkles the nose
AU10	Upper Lip Raiser	Raises the upper lip
AU12	Lip Corner Puller	Pulls corners of lips upward
AU14	Dimpler	Creates dimples by tightening lip corners
AU15	Lip Corner Depressor	Pulls corners of lips downward
AU17	Chin Raiser	Pushes chin boss and lower lip upward
AU20	Lip Stretcher	Pulls lips horizontally
AU23	Lip Tightener	Tightens and presses lips together
AU25	Lips Part	Parts lips
AU26	Jaw Drop	Lowers the jaw, parts lips
AU45	Blink	Closes and opens eyelids

Appendix B

PUBG Features

Table B.1: Player-Specific Features (Part 1)

Feature Name	**Description**
Health	
health	Current health level of the player
Combat	
total_knocks	Total number of enemies knocked down but not eliminated
total_kills	Total number of enemies eliminated by the player
total_armor_destructions	Number of times player has destroyed enemy armor
total_damage_dealt	Total damage inflicted to enemies
total_damage_taken	Total damage received by the player
headshot_ratio	Proportion of shots hitting enemy heads
torso_shot_ratio	Proportion of shots hitting enemy torsos
pelvis_shot_ratio	Proportion of shots hitting enemy pelvis area
leg_shot_ratio	Proportion of shots hitting enemy legs
penetration_shot_ratio	Proportion of shots that penetrated obstacles
total_multikills	Count of multiple eliminations in quick succession
accuracy	Overall shooting accuracy percentage
weapon_proficiency	Measure of player's skill with equipped weapons
max_kill_streak	Highest number of consecutive eliminations
combat_tempo	Rate of engagement in combat activities
throwable_items_used	Number of grenades and other throwables used
proximity_danger	Measure of nearby enemy threat level
Team Work	
total_revives_given	Number of teammates revived
total_assists	Number of eliminations assisted

Table B.2: Player-Specific Features (Part 2)

Feature Name	**Description**
Mobility	
total_on_foot_distance	Total distance traveled while on foot
total_distance_in_vehicle	Total distance traveled in vehicles
total_swimming_distance	Total distance traveled while swimming
Inventory Utilization	
boost_items_used	Number of boost/energy items consumed
heal_items_used	Number of healing items used
total_healed_amount	Total health points restored
Equipment Status	
headgear_level	Level of helmet/headgear equipped
vest_level	Level of body armor/vest equipped
backpack_level	Level of backpack equipped
main1_equipped	Primary weapon equipped
main2_equipped	Secondary weapon equipped
Looting Efficiency	
loot_box_items_looted	Number of items collected from standard loot boxes
carepackage_items_looted	Number of items collected from care packages
Zone Awareness	
distance_to_ safety_zone_edge	Distance to the edge of the current safe zone
cp_proximity	Distance to nearest care package

Table B.3: Game State and Match-Level Features

Feature Name	**Description**
Game State Features	
elapsed_time	Time passed since the match began
num_alive_players	Number of players still alive in the match
num_alive_teams	Number of teams with at least one living player
safety_zone_radius	Current radius of the safe play zone
Match-Level Features	
map_name	Name of the current map being played
team_size	Number of players per team (solo, duo, squad)
num_players	Total number of players in the match

Appendix C

Skill-Challenge Model Architecture and Hyperparameters

C.1 Model Architecture

The GCNSkillChallengeModel architecture is detailed below, showing the structure of each component and layer with their respective dimensions.

```
GCNSkillChallengeModel(
  (player_projection): Linear(in_features=51, out_features
     =128, bias=True)
  (game_projection): Linear(in_features=4, out_features=128,
     bias=True)
  (team_size_embedding): Embedding(3, 2)
  (map_embedding): Embedding(12, 6)
  (phase_embedding): Embedding(11, 4)
  (transformer): TransformerEncoder(
    (layers): ModuleList(
      (0-1): 2 x TransformerEncoderLayer(
        (self_attn): MultiheadAttention(
          (out_proj): NonDynamicallyQuantizableLinear(
             in_features=128, out_features=128, bias=True)
        )
        (linear1): Linear(in_features=128, out_features=256,
           bias=True)
        (dropout): Dropout(p=0.3, inplace=False)
        (linear2): Linear(in_features=256, out_features=128,
           bias=True)
        (norm1): LayerNorm((128,), eps=1e-05,
           elementwise_affine=True)
        (norm2): LayerNorm((128,), eps=1e-05,
           elementwise_affine=True)
        (dropout1): Dropout(p=0.3, inplace=False)
        (dropout2): Dropout(p=0.3, inplace=False)
      )
    )
  )
  (final_projection): Linear(in_features=140, out_features
     =128, bias=True)
```

```
  (skill_head): Sequential(
    (0): LayerNorm((128,), eps=1e-05, elementwise_affine=True
       )
    (1): Linear(in_features=128, out_features=64, bias=True)
    (2): GELU(approximate='none')
    (3): Linear(in_features=64, out_features=1, bias=True)
    (4): Sigmoid()
  )
  (challenge_head): Sequential(
    (0): LayerNorm((128,), eps=1e-05, elementwise_affine=True
       )
    (1): Linear(in_features=128, out_features=64, bias=True)
    (2): GELU(approximate='none')
    (3): Linear(in_features=64, out_features=1, bias=True)
    (4): ReLU()
  )
  (gcn): PlayerGCNEncoder(
    (player_proj): Linear(in_features=35, out_features=64,
       bias=True)
    (gcn_layers): ModuleList(
      (0): GCNLayer(
        (W): Linear(in_features=64, out_features=64, bias=
           True)
      )
      (1): GCNLayer(
        (W): Linear(in_features=64, out_features=16, bias=
           True)
      )
    )
  )
)
```

C.2 Training Hyperparameters

The model was trained with the following hyperparameters:

Hyperparameter	**Value**
Learning rate	3×10^{-4}
Weight decay	0.01
Warmup steps	$1,000$
Maximum epochs	100
Random seed	42
Batch size	128

Table C.1: Training hyperparameters for the GCNSkillChallengeModel

C.3 Model Components Description

The model architecture consists of the following key components:

- **Projection Layers:** Convert raw player features (51-dim) and game features (4-dim) into 128-dimensional embeddings.
- **Embedding Layers:** Encode categorical features like team size (3 categories to 2-dim), map type (12 categories to 6-dim), and game phase (11 categories to 4-dim).
- **Transformer Encoder:** Processes the sequence data with 2 transformer encoder layers using multi-head attention and feed-forward networks with dropout (p=0.3).
- **Graph Convolutional Network:** The PlayerGCNEncoder processes player interaction data through two GCN layers (64-dim and 16-dim).
- **Output Heads:** Two separate heads for predicting skill level (sigmoid activation) and challenge level (ReLU activation).

This architecture combines deep learning with graph neural networks to capture both sequential game dynamics and player interactions.