Self-avatar representation matters: Deciphering user immersion in VR games through Steam reviews

Dion Deng[1], Mila Bujić[1], Wangchi Lee[2], Mingrui Li[3], Juho Hamari[1]

[1] *Tampere University, Tampere 33100, Finland*
[2] *The Hong Kong University of Science and Technology, Hong Kong, China*
[3] *Hong Yi Cambridge International School, Changsha, China*

Abstract

This study critically examines the influence of self-avatars on user immersion in VR games by analyzing user reviews from Steam's top 100 VR games. Utilizing the BERT algorithm for text classification and detailed manual coding on avatar representations, the research addresses the effects of presence, perspective, visual features, and interactivity of avatars on immersion. Although the Mann-Whitney U test results were non-significant, effect size analyses revealed practical implications of avatar characteristics on user immersion. Notably, the study identifies key trends in avatar design within popular VR games, such as the predominance of first-person perspectives and the relative importance of hand representations over facial features. These findings suggest a need for a shift in focus in avatar research towards more user-relevant features. This innovative approach, using user-generated content, marks a significant departure from traditional experimental methods. It offers a richer, more ecologically valid understanding of user experiences in VR. The study's insights have significant implications for future avatar design and research.

Keywords

Avatar, virtual reality, product review, text classification, BERT, content analysis

1. Introduction

In immersive Virtual Reality (VR), self-avatars are the users' digital embodiment that play an important role in users' interaction and experience. The design and features of self-avatars are not merely aesthetic choices or tools to operate within the virtual environments, but also instrumental in determining the degree of immersion [1].

The immersion experience in VR is significantly influenced by the user's ability to identify with their avatar [2]. This identification is deeply rooted in the concept of presence, the sensation of being physically located in the virtual environment [2]. One important factor that is crucial to immersion experience is the similarity between the user's visual appearance and their avatar, which can encompass physical resemblance and behavioral and emotional congruence [3],[4].

In this context, understanding how different attributes of self-avatars, such as their presence, perspective, visual features, and interaction capabilities influence the user experience becomes paramount. Avatar research in VR has primarily relied on experimental methods, utilizing controlled lab environments to study user interactions and responses [5],[6],[7]. While these studies have provided valuable insights, their limited scope and controlled settings often fail to capture the diverse, real-world experiences of users. Such methods may not fully encompass the wide range of user backgrounds, preferences, and naturalistic behaviors that occur in everyday gaming contexts.

Our research adopts a novel approach by analyzing player feedback through game reviews in the context of avatars in VR. This method leverages the spontaneous, authentic, and varied opinions of the gaming community, providing a broader and more ecologically valid understanding of how self-avatars influence player experiences in real-life settings. User reviews, as a form of naturalistic data, offer insights into the aspects of self-avatars that resonate the most with players and significantly impact their sense of immersion and overall experience [8].

This study tries to explore the role of self-avatars in VR games through reviews, focusing on how their presence, perspective, visual features, and interactive capabilities impact user immersion. Our primary data source is user-generated reviews from the most popular VR games on the Steam platform for online distribution of games (Steam Inc). The methodology

8th International GamiFIN Conference 2024 (GamiFIN 2024), April 2-5, 2024, Ruka, Finland.

xiaohang.deng@tuni.fi (D. Deng); mila.bujic@tuni.fi (M. Bujić); juho.hamari@tuni.fi (J. Hamari)

0009-0006-1626-2365 (D. Deng); 0000-0002-4171-4806 (M. Bujić); 0000-0002-6573-588X (J. Hamari)

© 2024 Copyright for this paper by its authors. The use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

CEUR-WS.org/Vol-3669/paper12.pdf

involves a two-pronged approach: firstly, employing the BERT (Bidirectional Encoder Representations from Transformers) algorithm to identify and extract reviews specifically mentioning immersion; and secondly, utilizing a detailed codebook to manually code avatar features within these games. Through this innovative methodology, our study aims at analyzing a detailed picture of the current landscape of self-avatar design in VR games and how it aligns with users' experiences of immersion.

1.1. The effects of self-avatar in VR

Self-avatars not only serve as digital representations of players but also significantly influence their psychological experiences in VR [9],[10]. This section, grounded in extensive literature, examines how different aspects of self-avatar representations influence player immersion, aligning with our four leading research questions:

RQ1: How do presence of self-avatar and perspective affect the proportion of positive reviews about immersion of VR games?

RQ2: How do self-avatar's hand representation and body connectivity affect the proportion of positive reviews about immersion of VR games?

RQ3: How do visual features of self-avatar, including detail level, anthropomorphism, skin color, and body size affect the proportion of positive reviews about immersion of VR games?

RQ4: How does visual feedback of self-avatar interaction affect the proportion of positive reviews about immersion of VR games?

Presence and Perspective (RQ1): The incorporation of a self-avatar in VR is a fundamental element that significantly enhances the user's sense of embodiment, crucial for fostering a deep sense of presence and immersion within the virtual environment. Embodiment is central to VR experiences, directly influencing users' engagement and interaction within the virtual world [11],[12].

The choice of perspective, particularly between first-person and third-person views, further modulates this experience of embodiment. First-person perspectives are often associated with a higher sense of embodiment and presence, as they more closely mimic the natural human perception of their embodied perspective, offering a direct and uninterrupted view of the virtual environment from the avatar's eyes [13]. This immersive perspective allows users to directly align their physical movements with those of their avatars, creating a seamless and intuitive interaction that enhances the feeling of being in the virtual world [1].

In contrast, third-person perspectives provide a different type of interaction. While they offer a broader view of the avatar and its surroundings, they can sometimes create a sense of detachment, as the users view their avatars from an external standpoint [14]. However, this perspective can also be beneficial in certain scenarios, such as strategy games or situations where spatial awareness is key. Researchers have found that the choice of perspective can significantly affect how users process information and interact within VR [15].

Hand Representation and Body Connectivity (RQ2): Realistic hand representations can significantly boost the sense of agency and control, a key aspect of immersion. For example, when users see their virtual hands synchronized with their real movements, it enhances the sense of embodiment, resulting in a more engaging and intuitive VR experience [9].

The degree of body connectivity also impacts immersive experience. A fully connected avatar, as opposed to a disembodied hand or partial body representation, can increase the sense of bodily presence in the virtual environment [13]. This sense of a complete body in virtual space is crucial for a coherent experience, as it aligns with our natural perception of our bodies in the real world. The integration of proprioceptive feedback, where the user's movements are accurately reflected in the avatar, further enhances this sense of presence and embodiment [1].

Visual Features of Self-Avatars (RQ3): The visual features of self-avatars, including their level of detail, anthropomorphism, skin color, and body size, play a pivotal role in shaping user experiences in VR. These features significantly influence the degree of identification a user feels with their avatar, which in turn affects their immersion and overall experience.

The level of detail in an avatar's appearance can dramatically affect the user's sense of presence and immersion. High-resolution textures and detailed avatars can enhance the presence and engagement of the VR experience, leading to a stronger connection between the user and the virtual environment [4]. Detailed avatars enable users to identify more closely with virtual selves, fostering stronger embodiment.

Avatars with human-like features can enhance the social presence and emotional connection in VR, especially in multiplayer or social VR settings [16]. This connection can be particularly potent when avatars exhibit subtle human-like movements and micro-expressions.

The representation of diverse skin colors in avatars allows users from different backgrounds to find avatars that resemble and represent them, enhancing their sense of identity within the VR world [10]. The choice of an avatar's skin color can also impact the level of empathy and connection users feel with the virtual character.

The congruence of avatar size with the user's real body size can affect how users perceive spatial relationships and interact within the virtual reality environment [4]. This aspect is particularly important in applications where accurate spatial perception is crucial, such as in training simulations, modelling and other similar visualizations.

Our RQ3 explores the principles of Representation Theory [17], which posits that the effectiveness of information systems, such as VR environments, is significantly enhanced when they accurately represent real-life elements. By mirroring real-world characteristics, these avatars serve as authentic extensions of the user's identity within the virtual world. This fidelity in representation fosters a deeper connection and immersion, as users find it easier to

relate to and engage with avatars that closely resemble actual human features and behaviors.

Visual Feedback and Interaction in Self-Avatars (RQ4): The way self-avatars interact with the virtual environment and the corresponding visual feedback they provide are critical in shaping a user's immersion and overall experience in VR. This aspect of avatar design, encompassing the responsiveness and visual realism of avatar interactions, significantly contributes to the sense of presence and engagement within the virtual world [14].

Realistic interaction mechanics, such as accurate hand tracking and responsive movement, can deepen the user's sense of embodiment and agency within the virtual environment [1]. This visual realism aids in bridging the gap between the worlds, making the VR experience more intuitive and immersive.

The visual feedback from an avatar's actions, such as changes in the environment or reactions from other virtual entities, further amplifies the immersive experience. The feedback provides users with tangible consequences of their actions in VR, reinforcing the sense an active participant in the virtual world [12].

1.2. Game reviews as the data source

Avatar research has relied heavily on experimental methods to understand user behavior and experience. While these methods have been instrumental in advancing the field, they come with inherent limitations. Oulasvirta et al. (2016) [5] highlight that experimental settings often fail to replicate the complexity and variability of real-world scenarios, potentially leading to findings that lack ecological validity. Experiments typically involve small, non-representative samples, limiting the generalizability of the findings [6]. The controlled nature of these studies can also result in responses that do not fully capture the spontaneous and authentic reactions of users in naturalistic environments.

Recognizing these limitations, research on avatars has increasingly turned to alternative methods. One method is the analysis of user-generated content, such as user reviews. These reviews offer a rich, unfiltered, and authentic source of user feedback. Unlike the responses elicited in experimental settings, user reviews provide insights into the real-world experiences of a broad and diverse user base. This shift is supported by the growing understanding within the avatar studies that user experiences are multi-faceted and context-dependent [8].

Analyzing user reviews addresses several of the limitations inherent in experimental methods. It provides access to a diverse user sample, offering a level of representativeness that is often unattainable in lab-based studies. This is particularly important in VR, where user diversity significantly impacts interaction patterns and experiences [7]. As these reviews are generated in naturalistic settings, they offer a more accurate reflection of how users interact with and perceive technology in their daily lives, thus providing ecological validity that experimental studies might lack.

1.3. Proportion of reviews as a measure

Our methodological choice to quantify the proportion of user reviews addressing positive immersion experience as an indicator of the games' performance on immersion is underpinned by rigorous academic precedent. Quantitative content analysis of user-generated reviews is a well-established approach in the literature, which allows for the objective, systematic, and quantitative examination of communication content [18][19]. By focusing on the proportion of reviews that mention positive user experience of immersion, our study adopts a metric of salience that has been academically recognized as indicative of the importance or prominence of that topic within the consumer community [20].

This approach is grounded in the notion that the frequency of comments on a specific feature can be reflective of its significance to the user base, a methodological assumption that is supported by the Agenda-Setting Theory in mass communication [21]. This theory posits that the frequency of issues covered by the media influences the perceived importance of these issues among the public [21]. In the context of online reviews, the proportion of mentions can similarly set an 'agenda' by highlighting the features most impactful to users' experiences [22].

The reliance on proportional data is bolstered by research that suggests the volume and valence of mentions in reviews can act as proxies for consumer attitudes and satisfaction levels [23]. The significance of this method is further emphasized in [24],[25] which demonstrate a strong correlation between the proportion of review mentions of certain attributes and the consumer ratings of products.

By utilizing the proportion of topic mentions rather than the presence or absence of such mentions, we mitigate the risk of over-representing outlier opinions and instead capture a more balanced view of the collective sentiment. This is in line with previous findings[26]which highlight the robustness of proportional measures in depicting a more accurate reflection of the consensus among the user base.

Considering these theoretical and empirical foundations, our methodology is academically sound and provides a nuanced lens through which to assess the collective evaluation of a game's specific features by its users. The proportion of topic-specific reviews thus serves as a quantitative measure that is indicative of the overall perceived immersion. The method of conducting quantitative analysis with the proportion of positive reviews on immersion is introduced in the next section.

2. Methods

2.1. Data collection

From the Steam store, we narrowed the games by VR support (VR only) and language (English supported) to make sure our data is highly related to our research questions about VR and to avoid difficulties caused by multi-language text in the algorithm training. 2,938

games fulfilled the criteria. We selected the top 100 games as our sample is based on the number of user reviews, which gave us sufficient review data for the language model training.

To collect the reviews, we used Steam's official API, which provided data on all the reviews for games in STEAM, including the reviews' text, published date, time consumption on the game of the reviewers, etc. [7]. Our data collection was conducted on 25 Oct 2023, and a totally data of 282,847 reviews from 100 games was collected.

2.2. Data annotations

2.2.1. Game reviews

To train a text classification algorithm to detect reviews that reported positive user experience of immersion, we randomly selected 2,500 reviews (25 for each game) from our dataset as the training data.

We annotated the reviews related to positive user experience of immersion as 1, and others as 0 with the guide of a codebook (Table 1). In constructing this codebook, we aligned our inclusion and exclusion criteria with established games and theories, and prior empirical studies about immersion. Our purpose was to encompass a comprehensive range of elements that contribute to immersive experiences, as delineated by the following five aspects:

Terminology (Immersion and its related terms): Lombard and Ditton's (1997) [27] theory of presence emphasizes the psychological state where users feel immersed in a virtual environment. By focusing on direct references to 'immersion' or its synonyms, the codebook aligns with this theoretical framework, capturing users' perceived sense of being in the virtual world. For example, "The immersion in this VR game is really amazing; I totally forgot about the outside world" was included, and "I spent a lot of time immersed in this game," was excluded, where "immersed" refers to time spent, not describing the sense of immersion.

Specificity (Detailing specific experiences or emotions): Csikszentmihalyi's (1991) [28] concept of flow in gaming posits that immersion is often accompanied by detailed descriptions of experiences and emotions. This criterion ensures that the reviews analyzed are not just superficial mentions of immersion but reflect a deeper, flow-like engagement with the VR game. For example, we included "In this game, I totally felt the presence of being the character, as if I was really in that world.", and excluded "I really enjoy this game, it's fun to play," which is just a general experience share.

Game Design (Influence of game design elements on immersion): The Mechanics-Dynamics-Aesthetics (MDA) framework proposed by Hunicke, LeBlanc, and Zubek (2004) [29] illustrates how game mechanics influence player dynamics, including immersion. This criterion captures how users perceive and articulate the influence of game design on their immersive experiences. For example, we included "The 3D sound effects in the game made me feel like I was truly in another world" and excluded "The graphics of the game are beautiful," which did not mention how it affects immersion.

Interactivity (Enhancement of immersion through game's interactive features): The emphasis on interactivity aligns with Witmer and Singer's (1998) [30] Immersion Tendency Questionnaire, which suggests that interactive features of a game significantly contribute to the immersion experience. By coding for mentions of enhancement of immersion through interactivity, the codebook captures this aspect of the VR experience. For example, "The gesture control made me completely immersed in the game's actions" was included, and "The game controls are smooth," which did not mention immersion was excluded.

Real-World Comparison (Comparisons with real-world sensations): This criterion is based on the concept of 'Place Illusion' in VR [31], which argues that realistic, immersive VR experiences often lead to comparisons with real-world sensations. By coding for such comparisons, the codebook identifies instances where the immersive experience is strong enough to elicit real-world analogies, indicating a high level of presence. For example, we included "When I put on the VR headset, I completely forgot the outside world, as if I was in the game.", and excluded "This game made me forget my daily troubles," which did not specifically involve the immersive experience.

Finally, 146 (5.84%) reviews were annotated as 1 (related to the positive user experience of immersion).

Table 1
Annotation criteria of game reviews

Criteria	Inclusion	Exclusion
Terminology	direct references to "immersion" or its synonyms, indicating an explicit discussion of the immersive experience.	mention "immersion" only in a literal or non-contextual sense
Specificity	detailing specific experiences or emotions that convey a sense of immersion	general personal experiences not directly related to immersion
Game Design	analyze the influence of game design elements on immersion	general assessments of game design lacking a direct connection to immersion
Interactivity	emphasize the enhancement of immersion through the game's interactive features	focusing solely on interactivity without linking to immersion
Real-World Comparison	compare the gaming experience to real-world sensations to accentuate immersion's intensity	comparisons with the real world that do not specifically emphasize immersion

2.2.2. Game data

To classify avatar visual representations in the 100 VR games, we designed another codebook based on established games and theories, and prior empirical studies about avatar representation and embodiment. To observe self-avatar representations in each game, we used the keyword "game name + full gameplay" on YouTube and watched at least two of the gameplay videos with a minimum of five minutes on each video. After totally understanding every feature of the self-avatar representation, we annotated them based on the codebook (Table 2). For a further explication and descriptive data of the avatars' feature annotation, see section 3. Results.

In our annotation of avatars within the selected VR games, we paid particular attention to the aspect of personalization. For each feature of the avatar, such as skin color and body size, we assessed whether the game allowed players to personalize these elements. If a game offered the option for players to customize these aspects of their avatar, we labeled it as 'personalized'. Most of the games in our study do not have personalized self-avatars. Consequently, in our quantitative analysis, any data related to these customizable features were treated as missing due to insufficient data.

Table 2
Codebook of the avatar representations

Variables	Annotation rules
Presence	Visible or invisible avatar
First-person	First-person perspective support or not
Third-person	Third-person perspective support or not
Body type	Hands-only avatar or full-body/ upper-body avatar
Hand representation	Realistic or unrealistic hand representation
Hand accessories	With or without accessories on hands
Hand-body connection	Hand-body connects or not
Detail level	Include detailed textures or not
Anthropomorphism	With or without anthropomorphic features
Skin color (race)	Dark or light-skin
Skin color (transparency)	Skin color is semitransparent or non-transparent
Body size	Congruent avatar model size with human or incongruent (much bigger/smaller) avatar
Interactivity	Provide visual feedback caused by the avatar's interactivity or not

2.2.3. Reliability of annotation

To ensure the accuracy and consistency of our manual coding process, we conducted a preliminary annotation exercise with three independent coders. For the review data, three coders were tasked with analyzing a subset of 100 reviews, while for the avatar features, three coders each coded the characteristics of 10 games. Inter-rater reliability was evaluated using Cohen's k, which measures the level of agreement between coders beyond what would be expected by chance [32]. A score of .81 for the review data indicated good agreement, whereas for the avatar features a score of .68 suggested substantial agreement. Discrepancies in coding were reviewed in a series of consensus meetings where the coders discussed each disagreement until a unanimous decision was reached.

2.3. Topic detection

To classify the sentiment of reviews, we employed a state-of-the-art text classification algorithm BERT (Bidirectional Encoder Representations from Transformers) [33]. BERT is particularly well-suited for natural language processing tasks due to its deep learning architecture that considers the context from both the left and the right side of a token.

To train the model, we utilized a labeled dataset, where each review was pre-classified as either positive (1) or negative (0) based on the criteria in the codebook. The model was fine-tuned on this dataset, iterating through the corpus to learn the complex patterns associated with the sentiment expressed in gaming reviews.

We assessed the performance of our final BERT model using several evaluation metrics. The Receiver Operating Characteristic Area Under the Curve (ROC AUC) was 0.9679, indicating an excellent ability of the model to discriminate between the positive and negative classes. The ROC AUC is a performance measurement for classification problems at various threshold settings, where a score of 1 represents a perfect model and a score of 0.5 represents a model with no discriminative power.

In terms of precision, recall, and F1-score, which are critical metrics for classification problems, our model achieved the following results:

For class 0 (negative reviews), the model had a precision of 0.9916, meaning that 99.16% of the negative classifications were correct. The recall was 0.9834, indicating that 98.34% of the actual negative instances were correctly identified. The F1-score, a harmonic mean of precision and recall, was 0.9875.

For class 1 (positive reviews), the model achieved a precision of 0.7647 and a recall of 0.8667, resulting in an F1-score of 0.8125. This shows that while the model was slightly less precise in identifying positive reviews, it was robust in retrieving a high proportion of all relevant instances.

The overall accuracy of the model was 0.9766, demonstrating that it correctly classified 97.66% of the reviews. The macro average F1-score, which gives equal weight to both classes, was 0.9000, and the

weighted average F1-score, which accounts for class imbalance, was 0.9772.

3. Results

As the size is relatively small for our dataset, we selected the Mann-Whitney U test as a primary method of analysis. This non-parametric test is best suited to compare differences between two independent groups when the sample sizes are small, and distribution of data is not assumed normal. We also utilized Cliff's Delta as an effect size measure that gives a more relevant estimate of the scale of observed disparities in nonparametric situations. This approach supplements what we found with substantial practical understanding beyond just statistical significance. The results showed that some avatar features have impacts on perceived immersion. Especially, the realism of hand representation had a large effect size and thus it can be concluded that how hands represented in video game may considerably contribute to immersive experience. Contrastingly, detailed textures and skin color transparency reflected small to medium effect sizes showing that they had more modest impacts on immersion. The size of avatar and the provision of visual feedback from interactions also had medium effect sizes indicating their considerable role to improve immersive experience. These results note that certain avatar features may have different effects on the immersion of players in VR spaces. Table 3 summarizes the details for each feature in terms of statistics and effect sizes

Table 3
Summary of the data analyses

Variable	Median proportion of positive reviews on immersion in Group 1	Median proportion of positive reviews on immersion in Group 2	Mann-Whitney U	Asymp. Sig.	Effect size (Cliff's d)
Presence of Avatar	Visible 3.2% (N = 84, SD = .025)	Invisible 2.6% (N = 16, SD = .017)	765	.384	.138*
First-Person Perspective	Supported 2.7% (N = 92, SD = .025)	Not supported 3.7% (N = 8, SD = .017)	481	.153	.307**
Third-Person Perspective	Supported 2.5% (N = 19, SD = .178)	Not supported 3.0% (N = 81, SD = .253)	882	.325	.146*
Body Type	Hands-only 2.6% (N = 24, SD = .198)	Full-body/upper-body 2.5% (N = 61, SD = .269)	840	.294	.148*
Hand Realism	Realistic 2.4% (N = 49, SD = .025)	Unrealistic 2.6% (N = 34, SD = .026)	890	.068	.601***
Hand Accessories	With accessories 2.5% (N = 28, SD = .027)	Without accessories 2.4% (N = 29, SD = .022)	366	.528	-.099
Hand-Body Connectivity	Connected 2.3% (N = 18, SD = .020)	Not connected 2.6% (N = 65, SD = .027)	666	.374	.138*
Detail Level	Includes detailed textures 3.2% (N = 22, SD = .028)	Excludes detailed textures 2.5% (N = 63, SD = .024)	590	.304	-.149*
Anthropomorphism	With anthropomorphic features 2.8% (N = 75, SD = .023)	Without anthropomorphic features 3.3% (N = 25, SD = .028)	1037	.431	.106*
Skin Color - Light vs. Dark	Light skin 3.1% (N = 13, SD = .032)	Dark skin 2.1% (N = 6, SD = .013)	20	.106	-.487**
Skin Color - Transparency	Non-transparent 2.5% (N = 40, SD = .026)	Semi-transparent 2.0% (N = 4, SD = .024)	172	.403	.194*
Avatar Size	Congruent size 3.0% (N = 15, SD = .021)	Incongruent size 2.2% (N = 6, SD = .009)	26	.154	-.422**
Interaction Visual Feedback	With visual feedback 4.3% (N = 79, SD = .026)	Without visual feedback 2.5% (N = 6, SD = .018)	344	.068	.451**

Note. * small effect size (0.1 < |d| < 0.3), ** medium effect size (0.3 < |d| < 0.5), *** large effect size (|d| > 0.5).

4. Discussion

By analyzing user reviews from a selection of the most popular VR games on Steam, we found insights that challenge conventional understandings and potentially inspire novel perspectives in avatar research in the context of immersive virtual reality. Our discussions explore the multifaceted findings of our study, interpreting the implications of our results.

4.1. Overall results of the Mann-Whitney U and effect sizes

The predominance of non-significant results in our Mann-Whitney U initially appears to suggest a limited influence of avatar characteristics on player immersion. However, focusing on the effect sizes, rather than solely on statistical significance, offers a more nuanced understanding of our findings. Effect sizes provide insight into the magnitude of differences, and are less impacted by sample size, which is particularly informative in studies like ours where the sample size is relatively small [32]. However, it is important to note that the precision and confidence interval of these effect sizes is still influenced by the sample size.

These specific effect size results from our study carry important implications for future avatar research and design. The impact of avatar presence and perspective highlights the need for more targeted research to understand which aspects of avatar design resonate with different user demographics. For instance, future studies might explore how individual player characteristics, such as prior VR experience or personal preferences, interact with avatar features to affect immersion.

Moreover, our findings about the medium effect of first-person perspective on immersion suggest that VR game designers might consider offering players the option to choose their preferred perspective. This customization could cater to diverse player preferences, potentially enhancing the immersive experience for a broader user base.

Furthermore, the impact of skin color and avatar size in our study, which exhibited considerable effect sizes warrants special attention in design considerations. These factors were among the closest to reaching statistical significance, indicating their potential substantial influence on immersion in VR experiences. This suggests that even seemingly minor aspects of avatar design, like skin tone and body size, can have a profound impact on how users perceive and interact with the virtual environment.

4.2. Trends in avatar design for VR games

The descriptive analysis of the top 100 VR games on Steam provides a revealing glimpse into current trends in VR game design, particularly regarding avatar representation. A striking observation from our data is the scarcity of third-person perspectives, including indirect forms such as mirror reflections, in popular VR games. This trend has significant implications for the direction of avatar research.

Most avatar research has placed emphasis on visual attributes such as gender, age, and other identity markers. However, our findings suggest a disconnect between these research foci and the real-world VR games. In the absence of third-person perspectives or mirrors used in most popular VR games, features like facial appearance, gender, or age are less perceived. This raises questions about the relevance of such visual cues in first-person VR environments, where users primarily interact with the game world through their avatars' hands and actions.

Given this context, a shift in research focus appears necessary. Hand and lower-body representations in VR seem to be more critical for user immersion and interaction. This is supported by studies such as [9], which highlight the importance of hand representation in VR for enhancing the sense of control and embodiment. Additionally, research by [13] underscores the significance of embodiment in first-person VR experiences, further validating the need to focus on aspects directly experienced by the user.

Therefore, future avatar research might consider prioritizing the study of hand and full-body representations, exploring how their design, realism, and functionality contribute to immersion.

4.3. Avatar-user visual similarity and immersion

Our findings make a significant contribution to the understanding of avatar-user similarity and its impact on immersion in VR games. The nuances revealed through our analysis underscore the importance of similarity in fostering a deeper sense of immersion.

The preference for more detailed and realistic avatars aligns with the theory that higher fidelity in avatar design enhances the player's ability to relate to and identify with their virtual counterpart. This is supported by studies like that of [34], which found that users respond more positively to avatars that resemble their real-life appearance. Our findings extend this notion, suggesting that a detailed, high-fidelity avatar can act as an extension of the self within the virtual environment, thereby enhancing the immersive experience.

The observation regarding the preference for light-skinned avatars also points to a deeper aspect of user-avatar similarity but it could be reflective of the demographic composition of the VR gaming community. This phenomenon is echoed in the work [10], which demonstrated how skin color in avatars could influence the user's experiences and reactions in a virtual environment. Their results suggest that congruence in physical characteristics, such as skin color, between the avatar and the user can intensify the immersion, possibly due to enhanced identification.

Additionally, body size emerged as an influential aspect of avatar-user similarity. Our findings suggest that avatars with body sizes that closely match or are perceived as ideal by users can practically impact immersion. This is supported by research from [4], which demonstrated that the physical dimensions of

avatars, including their height and build, can affect the user's psychological responses in virtual interactions. An avatar with a relatable body size can create a more compelling and convincing representation of the user in the virtual world, contributing to a heightened sense of presence and immersion.

These insights suggest that VR developers should consider incorporating customizable avatars that can adapt to diverse user preferences, thereby enriching the overall user experience in virtual environments.

4.4. Methodological innovations and contributions to avatar studies

The methodology employed in this study represents a significant paradigm shift in avatar research, particularly in the field of VR. By harnessing the power of user-generated content in the form of Steam reviews, we have successfully introduced a novel approach to understanding how self-avatars impact user experiences in VR games. This method transcends the limitations of experimental research, offering a more authentic and comprehensive view of player perceptions and interactions with avatars.

Our approach, which combines the advanced natural language processing capabilities of BERT with meticulous manual coding, enables us to extract and analyze nuanced player feedback on a scale previously unachievable. This dual-method strategy effectively balances the need for large-scale data analysis with the subtlety of human interpretation, setting a new standard for research in this field. The use of BERT, particularly, exemplifies the cutting-edge of computational linguistics, offering unprecedented precision in identifying and classifying relevant user sentiments [33].

The significance of this methodology lies not just in its technical prowess but also in its ability to capture the diverse and multifaceted experiences of users. Unlike controlled experimental settings, our approach taps into a rich vein of real-world user interactions, encompassing a broad spectrum of opinions and experiences.

Furthermore, the insights garnered through this method offer invaluable implications for VR game design and avatar development. By understanding user preferences and perceptions as expressed organically in reviews, developers can tailor avatar designs more effectively to enhance user immersion and satisfaction. This user-centered approach to design is increasingly recognized as vital in creating engaging and impactful VR experiences [35].

5. Limitations and future research agenda

While our study provides valuable insights into self-avatar characteristics in VR games, it's crucial to acknowledge its limitations and outline potential avenues for future research.

5.1. Sample size and scope of analysis

The primary limitation of our study is the sample size (n=100), which constrains the depth and breadth of our analysis. With a larger dataset, more robust statistical methods like linear regression or factor analysis could be employed to uncover deeper insights [36]. The selected games are the most popular on Steam, which indicates that the findings in this study might not be generalizable to a broader range of VR experiences.

Expanding the study to include other VR platforms, such as Meta Quest, can provide a broader perspective. By analyzing user reviews across different platforms, researchers can capture a more diverse range of user experiences and preferences, as noted by [35] in their discussion on user experience research.

Future research should also consider longitudinal studies to track changes in user preferences and perceptions over time, as VR technology continues to evolve [37].

5.2. Sentiment classification

The initial plan to classify sentiment in immersion-related reviews encountered a limitation due to the small size of the training dataset. This small sample size restricts the robustness and generalizability of any sentiment classification model we could develop.

Despite the small dataset, our annotated data revealed a significant majority of positive immersion-related reviews (93.15%, n=136). This high proportion of positive sentiment is promising, suggesting that players generally perceive immersion aspects of VR games favorably. However, as [38] noted, sentiment analysis in complex domains like gaming can benefit from more nuanced classification, capturing a spectrum of sentiments rather than a binary positive/negative division.

For future research, expanding the dataset for training the sentiment model is crucial. A larger and more varied set of reviews would enable the development of a more sophisticated sentiment analysis model that can accurately differentiate between positive and negative sentiments regarding immersion.

Additionally, future studies should consider employing advanced machine learning techniques that can handle imbalanced datasets, as often seen in user-generated content where certain sentiments may dominate. Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) or ensemble learning methods can address the imbalance, as recommended by [39].

5.3. Observational limitations

The use of YouTube gameplay videos as a primary source for observing avatar characteristics in VR games presents several limitations that need consideration in future research.

While YouTube offers accessibility and a wide range of content, relying on gameplay videos for

detailed observations can lead to incomplete or skewed data. Gameplay videos are often edited and curated, potentially omitting crucial aspects of the gaming experience that are pertinent to avatar research. This limitation aligns with the concerns raised by [40], who note that online video content may not always represent the full spectrum of user experiences due to selective editing.

Another limitation is the potential for bias in the selection of videos. Content creators may have specific preferences or play styles that do not represent the average player's experience. This issue is highlighted by [41], who discusses how content creator biases can influence the portrayal of digital experiences in online videos.

To address these limitations, future studies could incorporate direct gameplay observation through platforms that offer unedited and comprehensive gameplay experiences. For instance, using data from beta testing sessions or developer-provided gameplay footage could yield more accurate and representative insights into avatar characteristics. Additionally, as suggested by [42], incorporating player interviews or surveys alongside gameplay observation can provide a more holistic understanding of player experiences and perceptions.

5.4. Impact of game type diversity

The diversity of game types in our sample of 100 VR games, ranging from RPGs to sports, music videos, and shooting games, introduces a potential confounding variable in our analysis of avatar representations. The variation in game genres can significantly influence how avatars are designed and interacted with, potentially impacting user perceptions of immersion.

The heterogeneous nature of game genres in our dataset could have diluted the specificity of our findings regarding avatar representations. As highlighted by [43], different game genres cater to different player expectations and experiences, which can significantly affect how players perceive and interact with avatars. For instance, the role and representation of avatars in an RPG might be fundamentally different from those in a sports game, leading to varied impacts on user immersion.

To address this issue, future research should consider focusing on specific game genres to control genre-related variance. This approach would allow for a more nuanced understanding of how avatar representations influence immersion within a particular gaming context. By isolating the variable of game type, researchers can more accurately assess the impact of avatars on the user experience.

Acknowledgements

This research was partially supported by the Academy of Finland (342144; 'POSTEMOTION').

References

[1] M. Slater, B. Spanlang, M. V. Sanchez-Vives, and O. Blanke, "First person experience of body transfer in virtual reality," PLoS One, vol. 5, no. 5, 2010, doi: 10.1371/journal.pone.0010564.

[2] M. V. Sanchez-Vives and M. Slater, "From presence to consciousness through virtual reality," Nature Reviews Neuroscience, vol. 6, no. 4. 2005. doi: 10.1038/nrn1651.

[3] J. N. Bailenson, J. Blascovich, A. C. Beall, and J. M. Loomis, "Equilibrium theory revisited: Mutual gaze and personal space in virtual environments," Presence: Teleoperators and Virtual Environments, vol. 10, no. 6, 2001, doi: 10.1162/105474601753272844.

[4] N. Yee and J. Bailenson, "The proteus effect: The effect of transformed self-representation on behavior," Hum Commun Res, vol. 33, no. 3, 2007, doi: 10.1111/j.1468-2958.2007.00299.x.

[5] A. Oulasvirta and K. Hornbæk, "HCI research as problem-solving," in Conference on Human Factors in Computing Systems - Proceedings, 2016. doi: 10.1145/2858036.2858283.

[6] M. Muller and S. Kogan, "Grounded theory method in hci and cscw," Cambridge: IBM Center for Social …, 2010.

[7] D. Deng, M. Bujic, and J. Hamari, "Understanding Multi-platform Social VR Consumer Opinions: A Case Study in VRChat Using Topics Modeling of Reviews," in Lecture Notes in Business Information Processing, 2023. doi: 10.1007/978-3-031-32302-7_4.

[8] C. Lu, X. Li, T. Nummenmaa, Z. Zhang, and J. Peltonen, "Patches and Player Community Perceptions: Analysis of No Man's Sky Steam Reviews," 2020.

[9] F. Argelaguet, L. Hoyet, M. Trico, and A. Lécuyer, "The role of interaction in virtual embodiment: Effects of the virtual hand representation," in Proceedings - IEEE Virtual Reality, 2016. doi: 10.1109/VR.2016.7504682.

[10] T. C. Peck, S. Seinfeld, S. M. Aglioti, and M. Slater, "Putting yourself in the skin of a black avatar reduces implicit racial bias," Conscious Cogn, vol. 22, no. 3, pp. 779–787, 2013, doi: 10.1016/j.concog.2013.04.016.

[11] M. Slater and S. Wilbur, "A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments," Presence: Teleoperators and Virtual Environments, vol. 6, no. 6, 1997, doi: 10.1162/pres.1997.6.6.603.

[12] F. Biocca, "The Cyborg's Dilemma: Progressive Embodiment in Virtual Environments [1]," Journal of Computer-Mediated Communication, vol. 3, no. 2, 2006, doi: 10.1111/j.1083-6101.1997.tb00070.x.

[13] K. Kilteni, R. Groten, and M. Slater, "The Sense of Embodiment in virtual reality," Presence: Teleoperators and Virtual Environments, vol. 21, no. 4. 2012. doi: 10.1162/PRES_a_00124.

[14] A. McMahan, "Immersion, engagement, and presence: A method for analyzing 3-d video

games," in The Video Game Theory Reader, 2013. doi: 10.4324/9780203700457-10.
[15] R. Tamborini and P. Skalski, "The role of presence in the experience of electronic games," in Playing Video Games: Motives, Responses, and Consequences, 2006. doi: 10.4324/9780203873700.
[16] S. J. Ahn, "Embodied Experiences in Immersive Virtual Environments: Effects on Pro-Environmental Attitude and Behavior," Dissertation:Stanford University, no. May, 2011.
[17] J. Burleson, V. Grover, J. B. Thatcher, and H. Sun, "A Representation Theory Perspective on the Repurposing of Personal Technologies for Work-Related Tasks," J Assoc Inf Syst, vol. 22, no. 6, 2021, doi: 10.17705/1jais.00707.
[18] M. Shelley and K. Krippendorff, "Content Analysis: An Introduction to its Methodology.," J Am Stat Assoc, vol. 79, no. 385, 1984, doi: 10.2307/2288384.
[19] K. A. Neuendorf, The Content Analysis Guidebook. 2020. doi: 10.4135/9781071802878.
[20] D. Riffe, S. Lacy, and F. Fico, Analyzing media messages: Using quantitative content analysis in research. 2014. doi: 10.4324/9780203551691.
[21] M. E. Mccombs and D. L. Shaw, "The agenda-setting function of mass media," Public Opin Q, vol. 36, no. 2, 1972, doi: 10.1086/267990.
[22] T. Tang, E. Fang, and F. Wang, "Is neutral really neutral? The effects of neutral user-generated content on product sales," J Mark, vol. 78, no. 4, 2014, doi: 10.1509/jm.13.0301.
[23] A. Nikolay, G. Anindya, and G. I. Panagiotis, "Deriving the pricing power of product features by mining consumer reviews," Management Science, vol. 57, no. 8. 2011. doi: 10.1287/mnsc.1110.1370.
[24] C. Vásquez, "Right now versus back then: Recency and remoteness as discursive resources in online reviews," Discourse, Context and Media, vol. 9, pp. 5–13, Sep. 2015, doi: 10.1016/j.dcm.2015.05.010.
[25] Y. Zhang, Y. Goh, and Q. Wang, "Unraveling the Effect of Competing Product Reviews on Consumer Choice and the Moderating Role of Consumer-Reviewer Peer Types," IEEE Trans Eng Manag, vol. 70, no. 10, p. 3315, 2023, doi: 10.1109/TEM.2021.
[26] R. Ullah, N. Amblee, W. Kim, and H. Lee, "From valence to emotions: Exploring the distribution of emotions in online product reviews," Decis Support Syst, vol. 81, pp. 41–53, Jan. 2016, doi: 10.1016/j.dss.2015.10.007.
[27] M. Lombard and T. Ditton, "At the heart of it all: The concept of presence," Journal of Computer-Mediated Communication, vol. 3, no. 2. 1997. doi: 10.1111/j.1083-6101.1997.tb00072.x.
[28] P. H. Mirvis and M. Csikszentmihalyi, "Flow: The Psychology of Optimal Experience," The Academy of Management Review, vol. 16, no. 3, 1991, doi: 10.2307/258925.
[29] R. Hunicke, M. Leblanc, and R. Zubek, "MDA: A formal approach to game design and game research," in AAAI Workshop - Technical Report, 2004.
[30] B. G. Witmer and M. J. Singer, "Measuring presence in virtual environments: A presence questionnaire," Presence: Teleoperators and Virtual Environments, vol. 7, no. 3, 1998, doi: 10.1162/105474698565686.
[31] M. Slater, D. Banakou, A. Beacco, J. Gallego, F. Macia-Varela, and R. Oliva, "A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality," Frontiers in Virtual Reality, vol. 3. 2022. doi: 10.3389/frvir.2022.914392.
[32] J. Cohen, "Statistical power analysis for the behavioural sciences. Hillside," NJ: Lawrence Earlbaum Associates. 1988.
[33] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in NAACL HLT 2019 - 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference, 2019.
[34] J. Bailenson, K. Patel, A. Nielsen, R. Bajscy, S. H. Jung, and G. Kurillo, "The effect of interactivity on learning physical actions in virtual reality," Media Psychol, vol. 11, no. 3, 2008, doi: 10.1080/15213260802285214.
[35] M. Hassenzahl and N. Tractinsky, "User experience - A research agenda," Behaviour and Information Technology, vol. 25, no. 2, 2006, doi: 10.1080/01449290500330331.
[36] Andy Field, Discovering Statistics using SPSS Statistics, vol. 66. 2009.
[37] S. Faisal, P. Cairns, and A. Blandford, "Building for users not for experts: Designing a visualization of the literature domain," in Proceedings of the International Conference on Information Visualisation, 2007. doi: 10.1109/IV.2007.32.
[38] B. Pang and L. Lee, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, 2008, doi: 10.1561/1500000011.
[39] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," Journal of Artificial Intelligence Research, vol. 16, 2002, doi: 10.1613/jair.953.
[40] M. Sjöblom and J. Hamari, "Why do people watch others play video games? An empirical study on the motivations of Twitch users," Comput Human Behav, vol. 75, 2017, doi: 10.1016/j.chb.2016.10.019.
[41] R. Rosalen, "YouTube: Online video and participatory culture," New Media Soc, vol. 21, no. 9, 2019, doi: 10.1177/1461444819859476.
[42] M. Consalvo and N. Dutton, "Game analysis: Developing a methodological toolkit for the qualitative study of games," Game Studies, vol. 6, no. 1, 2006.
[43] N. Wardrip-Fruin and P. Harrigan, First person: new media as story, performance, and game, vol. 42, no. 02. 2006. doi: 10.5860/choice.42-0713.