A Real-Time 3D Avatar System for Immersive Virtual Try-On on Public Displays

1st Yueqian Guo*
School of Design and Art
Jiangxi University of Finance and Economics
Nanchang, China
1202390090@jxufe.edu.cn

2nd Tianzhao Li*
School of Animation and Digital Arts
Communication University of China
Beijing, China
1826750732@qq.com

3rd Xin Lv†
School of Animation and Digital Arts
Communication University of China
Beijing, China
lvxincuc@163.com

*These authors contributed equally to this work.
†Corresponding author: Xin Lv (lvxincuc@163.com)

***Abstract*—Virtual Try-On (VTON) technology is emerging as a solution for public "mirror" displays, but interaction remains a key challenge. Touchscreens raise hygiene concerns, while mid-air gestures can suffer from fatigue and low social acceptability. This paper presents a real-time 3D avatar system designed to provide an immersive and practical VTON experience. The system's core contribution is the integration of a personalized 3D digital human with real-time motion capture, creating an intuitive "mirror" effect. A key feature is its touchless mid-air gesture framework, designed as a hygienic and immersive alternative to traditional inputs.**

To evaluate the system's effectiveness and user experience, we conducted a user study $(N = 20)$ comparing our optimized Gesture-Based system (A) against a Touch baseline (B) and a high-latency Gesture system (C). The results validated our system's design: the optimized Gesture system (A) achieved subjective usability (SUS) scores largely equivalent to the Touch baseline (B). Furthermore, our system (A) scored significantly highest on immersion and was rated highest for perceived hygiene.

Crucially, we found that physical fatigue was rated low for the optimized system (A), but high for the high-latency system (C). This demonstrates that user fatigue in gesture interfaces is primarily a product of system latency, not the physical act of gesturing. Our work confirms the viability of 3D avatar systems for public VTON, showing that a responsive gesture-based system can provide a highly immersive and hygienic experience that rivals the usability of touch.

***Index Terms*—Virtual Try-On (VTON), Human-Computer Interaction (HCI), Gesture Interaction, Digital Human, Usability Evaluation, Social Acceptability**

I. Introduction

The online fashion industry is experiencing remarkable growth. However, traditional online shopping has a significant obstacle: customers cannot physically try on garments [1]. This limitation creates uncertainty about product fit and style. This uncertainty can lead to low customer satisfaction [2].

To solve this problem, Virtual Try-On (VTON) technology has emerged. VTON systems allow users to visualize clothing on their bodies without wearing the items. These systems are increasingly deployed on large interactive displays in public or semi-public locations, such as retail stores [3]. Many of these applications function as a "virtual fitting room" or "mirror display" [4].

Despite their potential, the interaction methods for these public VTON systems present a significant Human-Computer Interaction (HCI) challenge. The choice of interaction modality directly impacts usability and user acceptance.

First, many systems on large vertical displays rely on direct touch input [3]. While familiar, this method has notable drawbacks in public settings. Users may have significant hygiene concerns about touching shared surfaces. Additionally, touch interaction requires the user to be close to the display [3]. This position can obstruct their view and break the intuitive "mirror" experience.

Mid-air hand gestures offer a touchless alternative. This interaction style is often considered more natural for 3D environments and immersive systems [5], [6]. However, gesture-only interfaces also have well-documented challenges. They can be imprecise for fine-grained selection [7]. Prolonged use can also lead to physical fatigue, sometimes called "gorilla arm." Moreover, performing large gestures in a public space can cause social awkwardness or embarrassment [8], [9].

This paper presents **a real-time 3D avatar system** designed to overcome these challenges and provide an immersive, practical VTON experience on public displays. The system's primary contribution is the integration of a personalized 3D digital human with a highly responsive, real-time motion capture pipeline [10]. This creates an intuitive "mirror" effect that is central to the user experience.

The key contributions of this work are:

1) **A Personalized Digital Human System:** We design and implement a VTON system based on a personalized 3D avatar. The system uses real-time motion capture to drive the avatar, providing users with an intuitive "mirror" experience [10].
2) **A Responsive Gesture Interaction Framework:** We develop and integrate a dedicated mid-air hand gesture framework [11]. This framework functions as a hygienic and immersive alternative to touch.
3) **A Controlled System Validation:** We conduct a user study $(N = 20)$ to evaluate the system's effectiveness. The study compares our optimized Gesture-Based system (A) against a Touch-Based baseline (B) and a high-latency Gesture system (C). We demonstrate that our responsive system (A) achieves usability scores equiva-

lent to touch, while scoring highest in **immersion** and **perceived hygiene**, and successfully mitigating **physical fatigue**.

II. Related work

A. Virtual Try-On (VTON) Technology

Virtual Try-On (VTON) systems allow users to visualize apparel on themselves without physically wearing it. Research in VTON generally follows two main approaches: 2D image-based methods and 3D model-based methods.

Image-based techniques are common. They typically superimpose a 2D garment image onto a user's portrait or video feed [12] [13] [14] [15]. A key challenge in this area is the accurate deformation, or warping, of the 2D garment to match the person's body shape and pose. Early methods often employed Thin-Plate Spline (TPS) transformations for this warping. More recent approaches utilize deep learning to estimate a dense appearance flow, which offers more complex and non-rigid warping

In contrast, 3D model-based approaches use 3D representations of both the user and the garments. These systems simulate the physical properties of the cloth on a 3D body model [2] [16] [17]. This allows for the simulation of cloth drape and fit, offering a more accurate and interactive experience, such as a 360-degree view. Our work builds upon this 3D avatar-based paradigm, focusing on its application for large public displays.

B. Interaction for Public Displays

Large interactive displays are increasingly common in public and semi-public locations, including shops, museums, and offices. The most prevalent interaction modality for these systems is direct touch [3] [18].

However, touch input presents significant challenges for large-scale public screens. First, on large vertical displays, portions of the screen may be beyond the user's physical reach. Second, in public settings, shared surfaces raise significant hygiene concerns. The COVID-19 pandemic, for example, accelerated the adoption of touchless alternatives due to concerns about disease transmission on shared touchscreens [19]. To overcome these limitations, researchers have explored other interaction methods, such as using external mobile devices or touchless body-based interaction, including mid-air gestures [20] .

C. Mid-Air Gesture Interaction

Mid-air gesture interaction is a prominent touchless alternative. Sensing technologies, such as depth cameras (e.g., Kinect) or optical trackers (e.g., Leap Motion), enable robust hand and body tracking [21] [11]. Many studies have demonstrated the utility of gestures for navigation, manipulation, and control in public settings [7] [6] [22].

Despite their benefits, mid-air gestures introduce their own usability challenges. Physical fatigue is a significant issue. Users must hold their arms aloft for extended periods, which can be tiring [23]. Another critical challenge is social acceptability. Users may feel self-conscious or embarrassed performing large or unusual gestures in a public space [24]. Therefore, effective gesture design must balance reliability, user comfort, and the social context.

D. Digital Human and Real-time Motion Capture

The concept of a "Digital Human" in the fashion context extends beyond a static avatar [8]. It is defined as a life-like, AI-powered being capable of interaction. In VTON systems, digital humans serve as personalized 3D models for realistic cloth simulation. A primary challenge is the creation and real-time animation of this digital human to mirror the user, which is essential for a public-screen "mirror" experience as described in our system.

Recent advancements in computer vision have enabled efficient, markerless motion capture from single RGB or depth cameras [25] [26]. These techniques can extract 2D or 3D skeletal data in real-time. This skeletal data can then drive the animation of a 3D digital human model. This process creates a "magic mirror" effect. Prior work shows that such mirror-like feedback enhances user engagement and immersion in interactive systems [4]. This markerless "sensor-to-skeleton" pipeline is foundational to our proposed real-time interaction system.

III. System design

This section details the architecture and core components of the proposed immersive virtual try-on (VTON) system. The system is designed for deployment on large-screen public displays, such as those in retail stores. The primary goal is to reduce the user's time and decision-making costs.

A. System Architecture

The system's core architecture is built upon four key modules. These modules process user input to deliver a real-time, interactive, and intelligent try-on experience.

The key modules are: Gesture Recognition Module and Motion Capture Module; Interaction Manager Core; Natural User Interface; Virtual Clothing Database and Digital Human Module;

The central data flow for the user interaction is as follows: user motion → capture / recognition → digital human movement / clothing switch / user interface feedback → large screen display.

B. Personalized Digital Human and Motion Capture

A realistic and responsive avatar is essential for an immersive "mirror" experience.

Avatar Generation: The system first creates a personalized digital human using Unreal Engine's MetaHuman [27], which allows us to create a digital human instances with a single picture. In the meantime, MetaHuman provides with various types of basic body shapes including male and female, tall, short and medium, Underweight, Medium and Overweight. Instead of requiring complex 3D scanners, the system uses

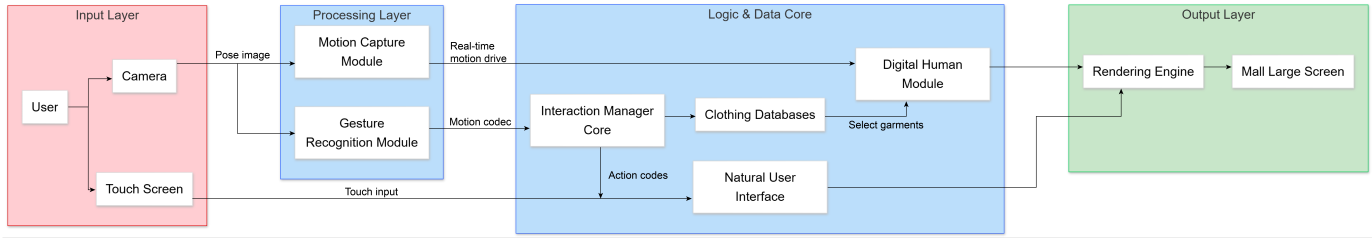

Fig. 1. System design overview

the built-in camera to capture the user's image and basic body shape. This process generates a digital human model that is visually similar to the user. This approach aligns with user preferences for mirror-like feedback, which has been shown to be highly effective in VTON user studies.

Motion Capture: To create an intuitive "mirror" effect, the system captures the user's movements in real-time. We use the Mediapipe framework [10] [26] for real-time pose estimation from the camera feed. This pose data drives the digital human's skeleton, allowing the avatar to replicate the user's actions instantly.

C. Gesture Interaction Control

Traditional VTON systems often rely on mouse or touch-screen inputs. This interaction method can limit the user's sense of immersion. As a touchless alternative, this system uses a Natural User Interface (NUI) framework based on mid-air hand gestures.

Gesture Control: Gesture recognition is managed through Mediapipe framework [26]. The system identifies key body pose information. This information is then processed by a "motion codec" to translate physical movements into discrete "action codes", allowing the user to browse and select garments. The specific action codes and their corresponding UI operations are as follows:

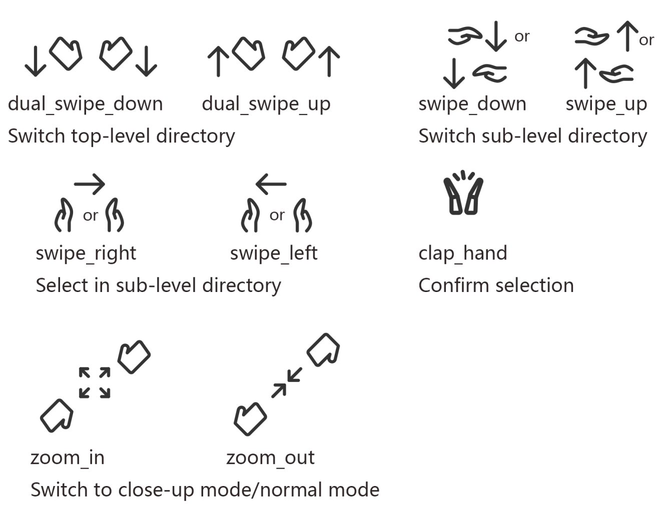

Fig. 2. Schematic diagram and actual meaning of "motion codec"

Interaction Logic: The system architecture is designed to accommodate both interaction modalities for our comparative experiment. The "Interaction Manager Core" processes inputs from either the "Gesture Recognition Module" or the "Touch Screen". Both action codes (from gesture) and touch commands are mapped to control the same UI elements (e.g., browsing the clothing database, selecting an item), ensuring a fair comparison of the two modalities.

Fig. 3. The specific user interface design

IV. Experimental Evaluation

We conducted a user study to evaluate the effectiveness and user experience of our proposed real-time 3D avatar system. The experiment was designed to validate our system's key design choices—particularly its low-latency gesture interaction—by comparing three conditions: (A) our optimized Gesture-Based system, (B) a Touch-Based baseline, and (C) a Gesture-Based system with intentional 0.5s latency.

A. Experimental Setup

Environment: The study was conducted in a lab environment. We simulated a retail shopping scenario. A large vertical display (80-inch) was used to replicate the "mirror" experience. Hardware: The system used a standard high-definition webcam running at 1080p/30fps for motion capture and image capture. The computer was equipped with intel 14700 and RTX 5070 to prevent performance bottleneck.

Participants: We recruited 20 participants (N=20). This sample size is consistent with formative usability studies in Human-Computer Interaction (HCI) and was determined sufficient to detect medium-to-large differences in usability

Fig. 4. A demonstration of experimental setup

metrics. Participants included 15 females and 5 males, aging from 18-25. The primary audience for VTON applications primarily consists of young females. However, our sample include a sufficient number of males to ensure the validation of results across a broader demographic. Participants were required to have experience with motion-controlled games but no significant experience with VTON systems (to avoid expert bias). Exclusion criteria included any individual unfamiliar with smartphone. preventing the use of hand gestures or touch screens, or a failure to complete all experimental tasks.

B. Experimental Tasks

Participants performed a series of typical virtual try-on tasks. The task set was designed to test core browsing, selection, and outfit coordination functions under both interaction conditions. The tasks included:

Task 1 (Browsing and Selection): Browse the "tops" category and successfully try on three different items.

Task 2 (Specific Search): Navigate the interface to find and try on a specific item (e.g., "one blue dress").

Task 3 (Outfit Coordination): Manually browse and select a matching item (e.g., "shoes" or "accessories") to complement the currently worn outfit.

C. Experimental Conditions

We used a within-subjects design. Each participant completed the full set of tasks under both conditions.

To mitigate learning effects, the presentation order of the conditions was counterbalanced. Participants were randomly assigned to one of two groups: Group 1 performed all tasks with Condition A (Gesture-Based) first, followed by Condition B (Touch-Based), then Condition C(Gesture-Based with delay). Group 2 performed the tasks in the reverse order.

Condition A (Gesture-Based System): Participants used the proposed VTON system. All interactions were performed using the mid-air hand gesture framework.

Condition B (Touch-Based Baseline): Participants used a baseline version of the system. All interactions were performed using a standard direct-touch interface on the large vertical display.

Condition C (Gesture-Based System with delay): Participants used the proposed VTON system, which was modified with delay of 0.5 second.

D. Measurement Metrics

We collected both objective and subjective data to evaluate the two conditions.

1) Objective Usability Metrics: We measured quantitative performance for each task:

Task Completion Time: The total time (in seconds) from the start of the task to its successful completion.

Operation Error Count: The number of incorrect operations. This included incorrect gestures (for Condition A) or failed touch inputs (for Condition B).

2) Subjective Evaluation Metrics: After completing the tasks in each condition, participants completed a questionnaire. The questionnaire was based on the Technology Acceptance Model (TAM) [28] and established HCI evaluation metrics. All items used a 5-point Likert scale. The order of the questionnaire questions is random to ensure that respondents are not influenced by the content of previous questions.

System Usability Scale (SUS): A standard 10-item questionnaire to measure overall system usability.

Perceived Ease of Use (PEOU): Measured the user's perception of interaction effort (e.g., "The system was easy to control").

Perceived Naturalness: Measured how intuitive the interaction modality felt (e.g., "The interaction felt natural").

Physical Fatigue: Measured the user's perceived physical strain (e.g., "Using this system was tiring").

Perceived Hygiene: Measured the user's comfort with the system's cleanliness in a public context.

Social Acceptability: Measured the user's willingness to use the system in the presence of others (e.g., "I would feel comfortable using this in a store").

Immersion: Measured the user's sense of engagement with the "mirror" experience.

V. Results and Analysis

A. Objective Usability Results

Box plots were used to visualize the objective performance metrics for all tasks.

For Task Completion Time, Condition B (Touch) was the fastest. Condition A (Gesture) consumed slightly more time(8%), but remained close to the baseline. Condition C (Degraded Gesture) cost significantly more time than Condition B(30%).

For Operation Error Count, Condition B (Touch) produced the fewest errors. In 3 specific tasks, Condition A (Gesture) resulted in 3.35 more errors than touch on average. Condition C (Degraded Gesture) had the highest error counts, which resulted in 10.35 more counts than Condition B on average.

These results suggest that while the proposed gesture interaction (A) is approaching the efficiency of mature touch interaction (B), a gap in operational accuracy still exists.

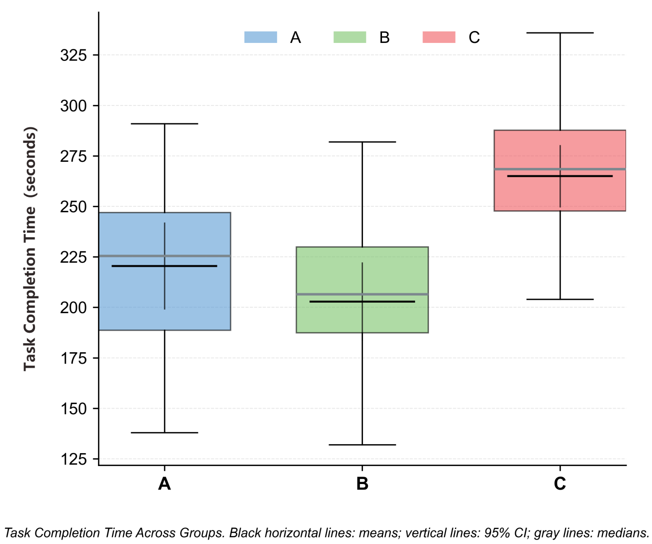

Fig. 5. Task Completion Time statistic results showed in box plot

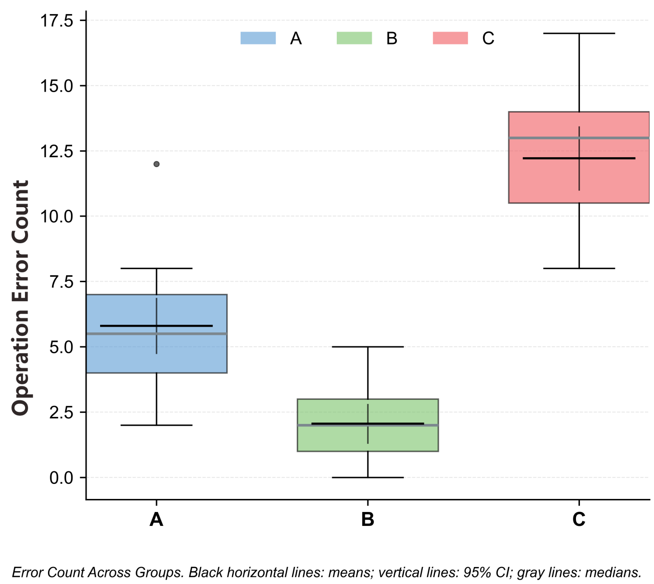

Fig. 6. Operation Error Count statistic results showed in box plot

B. Subjective Evaluation Results

Prior to analyzing the subjective data, a reliability and validity test was conducted on the questionnaire items (including SUS, PEOU, and other HCI metrics). The analysis yielded a **Cronbach's Alpha coefficient of 0.809**, indicating good internal consistency and reliability. The **Kaiser-Meyer-Olkin (KMO) value was 0.824**, suggesting the data was suitable for analysis [29].

1) Overall Usability (SUS): A comparison of the three conditions showed that Both Condition A (Gesture) and Condition B (Touch) demonstrated high usability scores. Condition A (Gesture) performed slightly inferior to Condition B (Touch) in overall usability. This parity suggests the emerging interaction style (A) performs favorably against the mature touch modality (B). Both Condition A and Condition B significantly outperformed Condition C (Degraded Gesture), highlighting the impact of system optimization.

2) Perceived Ease of Use (PEOU): For Perceived Ease of Use(Q16/17), Condition A (Gesture) showed a slight disadvantage compared to Condition B (Touch). However, Condition A performed better or equal to B on two of the four sub-items. Both A and B substantially outperformed Condition C (Degraded Gesture). This indicates that the perceived interactivity of the gesture system is approaching the usability of the mature touch baseline.

In terms of Perceived Naturalness(Q18), Condition A (Gesture) scored slightly lower than Condition B (Touch). Both A and B were rated as significantly more natural than Condition C (Degraded Gesture). The high familiarity of touch interaction, popularized by smartphones, likely contributed to its high naturalness rating.

For the indicator Immersion(Q19), Condition A (Gesture) scored the highest on immersion, followed by Condition B (Touch) and Condition C (Degraded Gesture). The gesture system (A) naturally tracked the user's body movements. Participants reported this created an intuitive "mirror-like" experience that enhanced immersion. In contrast, the high latency and poor recognition in Condition C broke this illusion and significantly reduced the sense of immersion.

C. HCI Specific Indicator Analysis

For the indicator Physical Fatigue(Q20), Both Condition A (Gesture) and Condition B (Touch) were rated low on physical fatigue. Participants did not find these systems tiring. In contrast, Condition C (Degraded Gesture) received high fatigue scores. This suggests that high system latency (C) increases both operation errors and the user's perceived physical strain.

For the indicator Perceived Hygiene(Q21), All conditions received relatively high hygiene scores. However, Condition B (Touch) scored slightly lower than the touchless conditions (A and C). It was observed that the touch display (B) showed fingerprints after use by multiple participants, which impacted this rating. The gesture-based modalities (A and C) demonstrated a natural advantage in perceived cleanliness.

For the indicator Social Acceptability(Q22), For social acceptability in a public setting, Condition B (Touch) was rated highest, followed by Condition A (Gesture), and finally Condition C (Degraded Gesture). This feedback suggests that participants felt some social inhibition when performing large, visible gestures in public. This effect was exacerbated in Condition C, likely due to the unresponsiveness and need for exaggerated movements.

VI. Conclusion and Discussion

This paper presented the design and implementation of a real-time 3D avatar system for immersive Virtual Try-On (VTON) on public displays. The system's design, which integrates a personalized digital human with a responsive, low-latency gesture framework, was validated through a controlled

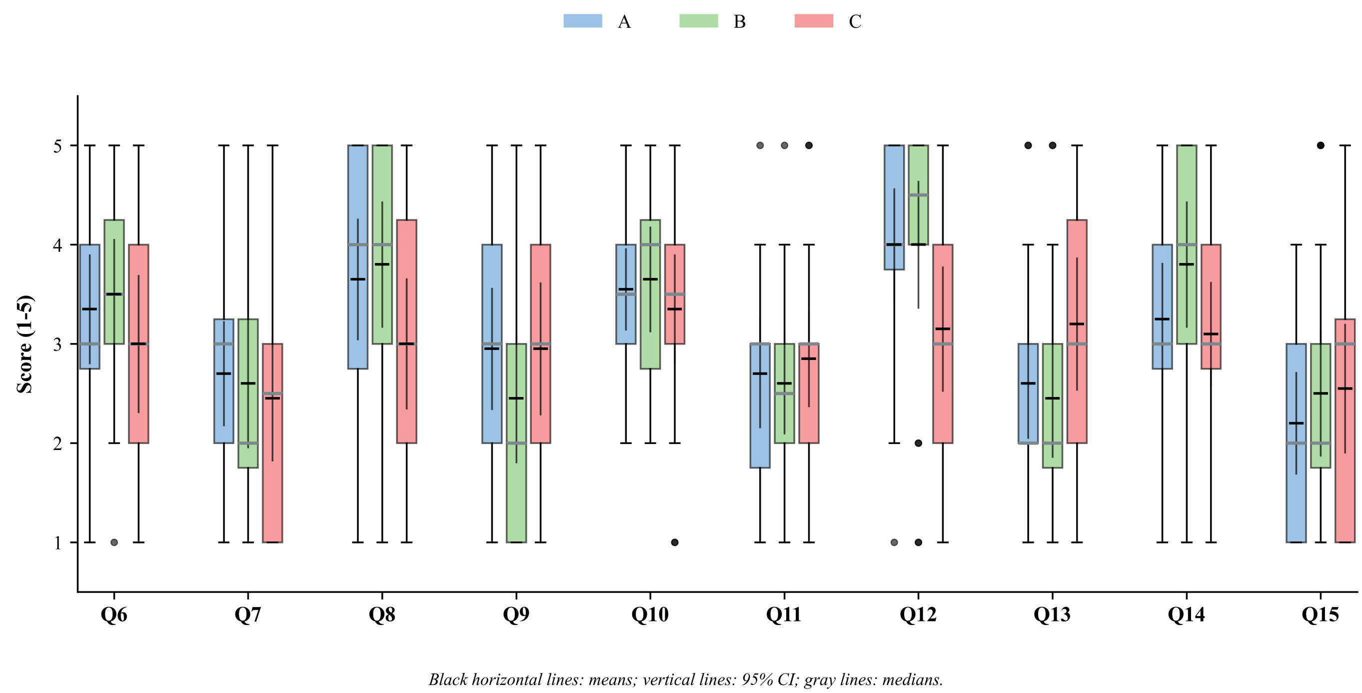

Fig. 7. SUS statistic results showed in box plot, Corresponding to Q6-Q15 in the questionnaire. For Q6/8/10/12/14/16, scores are positively correlated with performance. For Q7/9/11/13/15, scores are negatively correlated with performance.

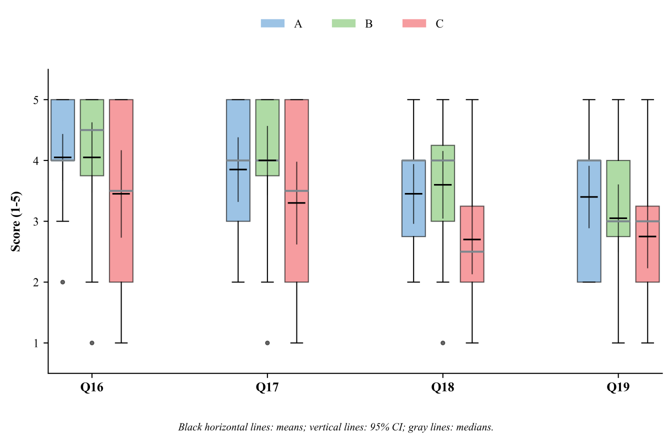

Fig. 8. PEOU statistic results showed in box plot, Q16/17 in the questionnaire was related to Perceived Ease of Use, Q18 was corresponded to Perceived Naturalness, Q19 was corresponded to Immersion. For Q16/17/18/19, scores are positively correlated with performance.

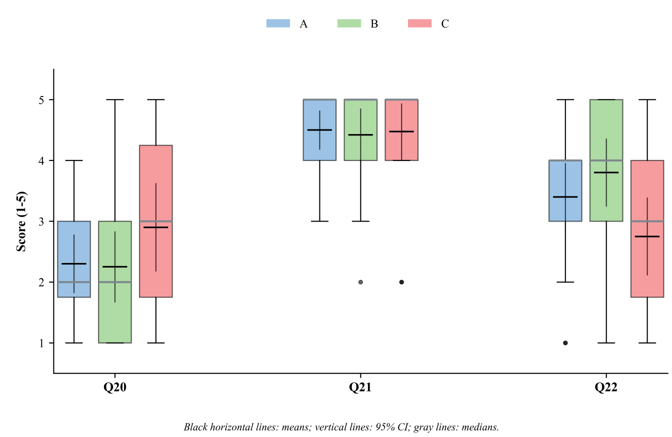

Fig. 9. HCI statistic results showed in box plot, Q20 was corresponded to Physical Fatigue, Q21 was corresponded to Perceived Hygiene, Q22 was corresponded to Social Acceptability. For Q21/22, scores are positively correlated with performance. For Q20, scores are negatively correlated with performance

user study $(N = 20)$. This study evaluated our optimized system (A) against a traditional touch baseline (B) and a high-latency gesture system (C).

A. Key Findings

Our experimental evaluation (Section V) confirmed the effectiveness of our system's design and provided key insights into its performance characteristics:

1) **System Responsiveness is Critical:** A primary finding is that **system latency, not the physical act of gesturing, is the primary cause of user fatigue**. Our optimized system (A) and the touch baseline (B) both received low physical fatigue scores (Fig. 9). In contrast, the high-latency system (C) received high fatigue scores. This validates our design emphasis on a responsive, low-latency motion capture pipeline.
2) **Achieving Experiential Advantages:** Our system (A) successfully provided a superior experiential quality. It was rated significantly **highest on immersion** (Fig. 8,

Q19), confirming that the "mirror-like" body tracking is highly effective for user engagement. This is a key design goal for VTON systems.

3) Practical Benefits: The touchless gesture framework of our system (A) demonstrated a clear, practical advantage in **perceived hygiene** (Fig. 9, Q21) over the touch baseline (B). This is a critical feature for systems deployed on public displays.
4) Usability: While the touch baseline (B) remains the most efficient modality (8% faster task completion, 3.35 fewer errors), the subjective usability (SUS) scores for our optimized system (A) were **largely equivalent** to touch (Fig. 7). This demonstrates that our system's gesture framework is viable and usable, achieving comparable usability to a mature, highly familiar interaction modality.
5) the Main Barrier: The primary challenge for our gesture-based system is **social acceptability** (Fig. 9, Q22). Participants felt more self-conscious performing gestures (A) in public than using the discreet touch interface (B).

B. Implications for Public VTON Interaction Design

Our findings offer several implications for the design of future 3D avatar systems:

- **Prioritize System Responsiveness:** The failure of Condition C confirms that latency is unacceptable. For gesture systems, minimizing processing delay is essential to prevent user fatigue, reduce errors, and maintain immersion.
- **Balance Immersion and Efficiency:** Designers must make a conscious choice. If the goal is rapid, high-throughput use (e.g., a simple kiosk), touch (B) is more efficient. If the goal is an engaging, branded "mirror" experience, the high immersion of a system like ours (A) may be more valuable.
- **Mitigate Social Awkwardness:** The low social acceptability of gestures must be addressed. Future system designs could explore more subtle "micro-gestures." Alternatively, designers could use physical enclosures, such as semi-private booths, to reduce social inhibition.
- **Leverage Hygienic Advantages:** The touchless nature of gesture-based systems is a strong and practical advantage. This benefit should be emphasized in the design and deployment of public systems.

C. Limitations and Future Work

This study has several limitations. First, the experiment was conducted with a limited sample size ($N = 20$) in a simulated lab environment. The social pressure in a real, crowded retail store is likely higher, which could further lower the social acceptability scores for gestures. Second, the experimental tasks were relatively simple. The performance gap between touch and gesture might widen with more complex interactions.

Future work will proceed in three directions. First, we plan to conduct a field study in a live retail environment to validate our system's performance. Second, we will explore and evaluate alternative gesture sets that are more subtle and socially acceptable. Third, we will continue to improve the VTON system itself, including enhancing the realism of the digital human [27] and expanding the virtual clothing database.

VII. Acknowledgment

We would like to express their gratitude to the Digital Human Research Institute of Communication University of China for its valuable support of this research. We also thank Google's Gemini service for its assistance in polishing the language and improving the clarity of this paper.

References

[1] T. Islam, A. Miron, X. Liu, and Y. Li, "Deep learning in virtual try-on: A comprehensive survey," *IEEE Access*, vol. 12, pp. 29 475–29 502, 2024.

[2] G. R.R.N.P.A.B.W.M.S.R, H. Wickramarathna, G. E.H.N.L., K. M. Purnima, S. Reyal, and S. Siriwardana, "Elegant fit-on – virtual fitting room on handheld devices," in *2022 4th International Conference on Advancements in Computing (ICAC)*. IEEE, 2022, pp. 180–185.

[3] C. Ardito, P. Buono, M. F. Costabile, and G. Desolda, "Interaction with large displays: A survey," *ACM Comput. Surv.*, vol. 47, no. 3, feb 2015.

[4] M. Yuan, I. R. Khan, F. Farbiz, S. Yao, A. Niswar, and M.-H. Foo, "A mixed reality virtual clothes try-on system," *IEEE Transactions on Multimedia*, vol. 15, no. 8, pp. 1958–1968, December 2013.

[5] A. Nassani, G. S. V. P. Raju, and S. M. Sameer, "A review of gestural interaction techniques in virtual environments," *International Journal of Computer Science and Engineering (IJCSE)*, vol. 3, no. 2, pp. 59–65, 2014.

[6] V. S. Souza, A. G. G. D. S. Conte, L. P. C. T. da Silva, A. C. de Paiva, C. A. Vilupar, and E. W. G. Clua, "A review of interaction techniques for immersive environments," *IEEE Access*, vol. 10, pp. 70 138–70 162, 2022.

[7] P. Koutsabasis and P. Vogiatzidakis, "Empirical research in mid-air interaction: A systematic review," *International Journal of Human-Computer Interaction*, feb 2019.

[8] S. B. Far and A. I. Rad, "Applying digital twins in metaverse: User interface, security and privacy challenges," *Journal of Metaverse*, vol. 2, no. 1, pp. 8–15, jun 2022.

[9] D. Medeiros, M. McGill, A. Ng, R. McDermid, N. Pantidi, J. Williamson, and S. Brewster, "From shielding to avoidance: Passenger augmented reality and the layout of virtual displays for productivity in shared transit," *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 11, pp. 3640–3650, 2022.

[10] J.-W. Kim, J.-Y. Choi, E.-J. Ha, and J.-H. Choi, "Human pose estimation using mediapipe pose and optimization method based on a humanoid model," *Applied sciences*, vol. 13, no. 4, p. 2700, 2023.

[11] C. Groenewald, C. Anslow, J. Islam, C. Rooney, P. Passmore, and W. Wong, "Understanding 3d mid-air hand gestures with interactive surfaces and displays: A systematic literature review," in *Proceedings of British HCI 2016 Conference Fusion*. BCS Learning and Development Ltd., 2016.

[12] S. He, Y.-Z. Song, and T. Xiang, "Style-based global appearance flow for virtual try-on," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 3470–3479.

[13] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis, "Viton: An image-based virtual try-on network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7543–7552.

[14] S. Choi, S. Park, M. Lee, and J. Choo, "Viton-hd: High-resolution virtual try-on via misalignment-aware normalization," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 14 131–14 140.

[15] Z. Xie, Z. Huang, X. Dong, F. Zhao, H. Dong, X. Zhang, F. Zhu, and X. Liang, "Gp-vton: Towards general purpose virtual try-on via collaborative local-flow global-parsing learning," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 23 550–23 559.

[16] Z. He, Y. Ning, Y. Qin, G. Wang, S. Yang, L. Lin, and G. Li, "Vton 360: High-fidelity virtual try-on from any viewing direction," in *Proceedings of the Computer Vision and Pattern Recognition Conference*, 2025, pp. 26 388–26 398.

[17] F. Zhao, Z. Xie, M. Kampffmeyer, H. Dong, S. Han, T. Zheng, T. Zhang, and X. Liang, "M3d-vton: A monocular-to-3d virtual try-on network," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13 239–13 249.
[18] D. Vogel and R. Balakrishnan, "Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users," in *Proceedings of the 17th annual ACM symposium on User interface software and technology*, 2004, pp. 137–146.
[19] V. Mäkelä, J. Winter, J. Schwab, M. Koch, and F. Alt, "Pandemic displays: Considering hygiene on public touchscreens in the post-pandemic era," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–12.
[20] C.-M. Iacob, A. Uusberg, R.-H. Tunnel, J. K. Kaasik, M. Vasser, J. Aru, C. Sequeros, H.-M. Pedastsaar, Botha-Ravy, D. Eisinger, H. P. H. Shum, H. Wang, T. Komura, and D. Cohen-Or, "Eliciting multimodal and collaborative interactions for data exploration on large vertical displays," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2022, pp. 386–395.
[21] D. Yu, Q. Zhou, T. Dingler, E. Velloso, and J. Goncalves, "Blending on-body and mid-air interaction in virtual reality," in *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2022, inferred from metadata and code repository link [2963].
[22] A. M. Bernardos, X. Wang, L. Bergesio, J. A. Besada, and J. R. Casar, "Assessing the acceptance of a mid-air gesture syntax for smart space interaction: An empirical study," *Journal of Sensor and Actuator Networks*, vol. 13, no. 2, p. 25, 2024.
[23] F. Ma, F. Song, Y. Liu, and J. Niu, "Quantitative analysis on the interaction fatigue of natural gestures," *IEEE Access*, vol. 8, pp. 190 797–190 811, 2020.
[24] M. Koelle, S. Ananthanarayan, and S. Boll, "Social acceptability in hci: A survey of methods, measures, and design strategies," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–19.
[25] S. Qiao, Y. Wang, and J. Li, "Real-time human gesture grading based on openpose," in *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*. IEEE, 2017, pp. 1–6.
[26] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. G. Yong, J. Lee *et al.*, "Mediapipe: A framework for building perception pipelines," *arXiv preprint arXiv:1906.08172*, 2019.
[27] Z. Fang, L. Cai, and G. Wang, "Metahuman creator the starting point of the metaverse," in *2021 International Symposium on Computer Technology and Information Science (ISCTIS)*. IEEE, 2021, pp. 154–157.
[28] Y. Toraman and B. B. Geçit, "User acceptance of metaverse: an analysis for e-commerce in the framework of technology acceptance model (tam)," *Sosyoekonomi*, vol. 31, no. 55, pp. 85–104, 2023.
[29] M. E. Swerdlik and R. J. Cohen, "Psychological testing and assessment: An introduction to tests and measurement," *Journal of Disability Policy Studies*, vol. 22, no. 2, pp. 83–94, 2005.