Perceptual Sensitivity to LOD-Induced Artefacts in Animated Personalised versus Non-Personalised Avatars

Guilherme Noronha Jardim
Warwick Manufacturing Group
University of Warwick

Supervised by David Petrescu

1 September 2025

WARWICK
THE UNIVERSITY OF WARWICK

Abstract

As real-time graphics technology improves, maintaining high-quality realism in virtual humans becomes increasingly demanding on computing resources. To address this, Level of Detail (LOD) systems are widely used, adjusting the visual complexity of avatars depending on factors like camera distance or system performance. Despite their importance, we still do not fully understand how discrete, geometry-based abrupt LOD popping affect user perception, especially in cases where users feel a sense of ownership or identification with an avatar.

This study examines how easily users notice changes in LOD during live animations featuring digital representations of humans with MetaHuman technology. It compares whether users are more sensitive to these visual changes when viewing neutral avatars or customised avatars that resemble themselves, given prior evidence that similarity/ownership and viewing context can modulate perception. A fixed-reference Method of Constant Stimuli was employed with abrupt LOD pop between reference model and test levels (ranging from the full-resolution mesh to reductions of over 95% for the body and ≈99% for the face). Participants indicate the moment they detect a change via mouse click. Analyses estimated psychometric detection thresholds via logistic modelling and compared thresholds across avatar conditions.

Detection crossed 50% around mid-range simplifications. Motion increased detectability for neutral avatars but not for customised ones, and personalisation did not lower thresholds as hypothesised.

By targeting dynamic, non-VR animations, this work extends prior LOD and realism research that has typically focused on static stimuli or VR/2D contexts. The outcome is practical guidance for allocating fidelity budgets in digital humans so that developers can balance perceptual quality against performance where it matters most.

Contents

Glossary 7

1 Introduction 9

1.1 Technical burden of high-fidelity avatars 9

1.2 Known perceptual side-effects of LOD transitions 10

1.3 Why ownership and similarity-based factors might amplify sensitivity . 12

1.4 Identified Research Gap 13

1.5 Research Questions and Objectives 15

1.6 Thesis roadmap 16

2 What makes an avatar look 'right'? 17

2.1 Visual realism, the uncanny valley and fidelity trade-offs 17

2.2 Similarity, ownership and the role of perspective 19

2.3 Takeaways for LOD design 20

3 How motion governs perception 21

3.1 Motion-dominant cue findings 21

3.2 When motion meets appearance 22

3.3 Summary 25

4 Perceptual LOD thresholds and resource budgets **26**

4.1 Geometry-based thresholds (mesh simplification) 26

4.2 Rendering-based thresholds (Gaussian avatars & facial detail) 27

4.3 Context moderators: distance, motion complexity and socio-emotional salience 27

4.4 Implications for LOD budgets 29

5 Methodology **30**

5.1 Overview and rationale 30

5.2 Participants 30

5.3 Design 32

5.4 Hypotheses 32

5.5 Apparatus and Stimuli 33

5.5.1 Software 33

5.5.2 Hardware and display 33

5.5.3 Viewing conditions 33

5.5.4 Avatars 34

5.5.5 Animations 36

5.5.6 LOD control 36

5.5.7 Visual context and camera 37

5.5.8 Timing and trial generation (summary) 37

5.6 Procedure 38

5.6.1 Session Flow 38

5.6.2 Instructions 38

5.6.3 Warm-up (excluded from analysis) 39

5.6.4 Trial structure 39

5.6.5 Randomisation and counterbalancing 39

5.6.6 On-screen display and feedback 39

5.7 Data logging and quality control 40
5.7.1 Logged fields (segment-level schema) 40
5.7.2 Primary scoring (per-segment) 41
5.7.3 Quality control rules 41
5.7.4 Derived analysis dataset 42
5.7.5 Planned exclusions in modelling 42
5.7.6 Reproducibility 42
5.8 Data preprocessing 43
5.9 Analysis plan 44
5.9.1 Primary analyses 44
5.9.2 Justification of trial method and modelling choices. 44
5.9.3 Group summaries 45
5.10 Deviations and contingencies 45

6 Results 47
6.1 Sample & data quality 47
6.2 Descriptives 49
6.2.1 Psychometric fits 51
6.3 Model diagnostics & robustness 52
6.4 Interim summary 52

7 Discussion 54
7.1 Recap & positioning 54
7.2 Principal findings 55
7.3 Interpretation 56
7.3.1 Tolerance to degradation for Customised avatars. 57
7.3.2 Why motion helps Neutral but not Customised. 58
7.3.3 What the task setup contributes. 58

7.3.4 Sensitivity or criterion (or both). 59

7.3.5 What the results do *not* show 59

7.3.6 Why this still matters for practice. 59

7.3.7 Summary 60

7.4 Practical implications for real-time characters 60

7.4.1 Budget the LOD envelope per shot 60

7.4.2 Motion-aware switching 61

7.4.3 Where to spend triangles 61

7.4.4 Design rules for the switch itself 61

7.4.5 Screen-space and camera pragmatics 61

7.4.6 Instrumentation and QA 62

7.4.7 Production checklist (condensed) 62

7.5 Limitations and threats to validity 62

7.5.1 Sample size and precision 63

7.5.2 Model choices and fallbacks 63

7.5.3 Stimulus scope and presentation 63

7.5.4 Task design and measurement 63

7.5.5 Personalisation construct 64

7.5.6 Carry-over, learning and fatigue 64

7.5.7 Data quality and logging constraints 64

7.5.8 Population and context 64

7.5.9 Summary 65

7.6 Future work 65

7.6.1 Power and participant diversity 65

7.6.2 Modality factors and generalisation 65

7.6.3 Task design and modelling 66

7.6.4 Ownership and personalisation measurement 66

7.6.5 Transition design and masking. 66
7.6.6 Deployment and telemetry 66
7.6.7 Multi-character and social contexts 67
7.7 Summary 67

8 Conclusion **68**
8.1 Summary of aims and findings 68
8.2 Key insights and contributions 69
8.2.1 A task-oriented taxonomy that positions the study. 69
8.2.2 Operational, engine-mapped thresholds. 69
8.2.3 Avatar-dependent role of motion. 70
8.2.4 A scoring/QC pattern that travels. 71
8.2.5 What this buys for production. 71
8.3 Final remarks 71

Appendices **73**
.1 Ethics approval 74
.2 Consent Form 78

Glossary

Avatar (Neutral) The default, non-personalised MetaHuman used as a baseline stimulus.

Avatar (Customised) A MetaHuman adjusted to resemble the participant (face/clothes/skin tones etc.), used to test whether personalisation changes sensitivity to artefacts.

Level of Detail (LOD) A discrete model fidelity level (here, integers 0-7) selected by the engine according to distance or performance heuristics. In this experiment, higher LOD indices denote lower geometric detail

GeoLOD Geometry-only LOD in which mesh complexity (e.g., vertices/triangles) is reduced while materials, lighting and animation remain constant.

LOD pop The visible, abrupt change when the renderer switches between LODs. This thesis measures how detectable such changes are.

Reference segment (LOD 0) The baseline interval in each trial shown at the highest fidelity (LOD 0).

Test segment (LOD X) The interval within a trial in which the avatar temporarily switches from LOD 0 to a test level $X \in \{1, \ldots, 7\}$.

Trial (0$\rightarrow X \rightarrow$0) One complete presentation consisting of a Reference segment, a Test segment at LOD X, and a return to the Reference.

Detection click The participant's button-press indicating they detected an LOD change during the Test segment.

Hit / Miss A *hit* is a detection click within the Test segment; a *miss* is no click during the Test segment.

Inconsistency click A click outside the Test segment (e.g., during Reference or the return interval). Used here as a task-consistency check rather than a classic "false alarm".

Reaction time (RT) Time (ms) from Test-segment onset to the detection click.

Anticipation filter Quality-control rule discarding unrealistically fast responses (here, RT < 150 ms).

Method of Constant Stimuli (MoCS) A psychophysical design presenting a fixed set of stimulus levels (here, LODs 1-7) multiple times in randomised order to estimate a detection function.

Psychometric function The function mapping stimulus level to detection probability. Unless stated otherwise we fit a logistic (sigmoid):

$$p(\text{detect} \mid X) \;=\; \frac{1}{1+\exp[-(\beta_0+\beta_1 X)]}.$$

Detection threshold $x_{0.5}$ The LOD level at which the fitted psychometric function reaches 50% detection. For the logistic above, $x_{0.5} = -\beta_0/\beta_1$.

Slope β_1 The steepness of the psychometric function; larger $|\beta_1|$ indicates faster growth in detection probability across LODs.

Social Presence The subjective feeling of being "with" another sentient agent in a mediated setting.

Pooling Aggregating data across participants or conditions (e.g., to plot a single psychometric curve per condition), typically after computing per-participant proportions.

Counterbalancing (AB/BA) Alternating the order of avatar blocks (Neutral/Customised) across participants to control for order effects.

Viewing conditions Screen, distance and framing used to standardise retinal size and viewpoint across participants.

Return validation QC step ensuring each retained response cleanly corresponds to one Test segment (e.g., removing stray/duplicate clicks spanning the return to Reference).

Chapter 1

Introduction

1.1 Technical burden of high-fidelity avatars

Developers face the challenge of balancing visual realism with computer limitations as virtual environments increasingly utilise high-fidelity digital people (Lombardi et al., 2018). Rendering highly detailed avatars continuously places substantial demands on computational resources, and a common approach to managing this is through Level of Detail (LOD) systems. These systems dynamically adjust the visual complexity of avatars depending on factors such as camera distance, scene complexity, or hardware constraints.

In modern engines, LOD typically encompasses multiple sub-systems: mesh decimation (reduced vertex/triangle counts), simplified deformation rigs and facial controllers, texture resolution and streaming budgets, hair/groom cards, and occasionally shader/lighting approximations. While these changes are often triggered by distance heuristics, real-time cinematics or performance-sensitive sequences may also trigger LOD changes based on instantaneous GPU/CPU load or scene complexity. The practical question is not *whether* to use LOD, but *how aggressively* a developer can reduce fidelity before perceptual costs outweigh performance gains. In this work, only geometry-based LOD (GeoLOD) is manipulated, treating fidelity as a perceptual budget to be allocated where it is most salient.

MetaHuman avatars provide eight discrete geometry LODs (see Figure 5.3), ranging from full detail (LOD 0) to heavily simplified meshes (LOD 7).

Figure 1.1 illustrates a general example of mesh decimation under different algorithms, highlighting how resolution can be reduced while either preserving or degrading structural regularity. This provides a broader context for the specific GeoLOD manipulations applied later in this study.

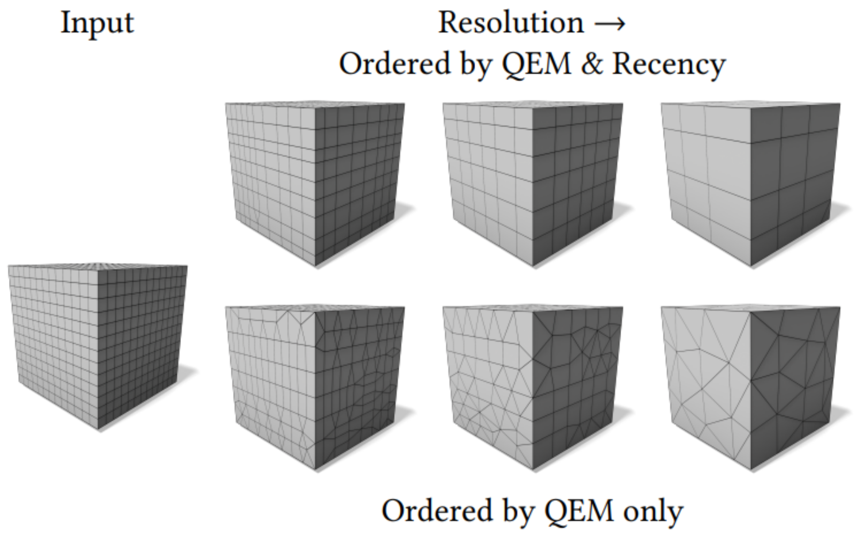


Figure 1.1: General mesh decimation across resolutions. Top row: simplification ordered by Quadric Error Metric (QEM) and recency, preserving quad structure. Bottom row: simplification ordered by QEM only, producing irregular triangulation. After Knodt (2025), Fig. 5.

1.2 Known perceptual side-effects of LOD transitions

Changes in LOD, particularly abrupt transitions, may negatively affect users' perception of realism, potentially disrupting the immersive experience (Kenny et al., 2019; Hodgins et al., 2010). Prior research demonstrates that viewers are sensitive to inconsistencies or anomalies in animated avatars when those avatars attempt realistic human-like motion, compared with stylised or mechanically simplified motions (Hodgins et al., 2010). Moreover, studies have shown that people perceive animations differently depending on their visual fidelity, indicating that avatars do not necessarily need the highest level of realism to achieve high avatar perception results (Lee et al., 2025). For example, research by Kang et al. (2022) demonstrated that mid-level avatar facial detail in augmented reality (AR) could achieve similar levels of social presence to highly detailed avatars, significantly reducing computational demand without noticeably affecting perception. Additionally, Nguyen et al. (2024) showed that substantial vertex simplifications of up to 95% could remain imperceptible to users at moderate viewing distances, offering practical thresholds for dynamic LOD systems in Virtual Reality (VR). Crucially, motion cues dominate user perception of realism and physicality, often more so than surface detail, especially in scenarios involving conversation or action perception (Yamac et al., 2024; Ennis et al., 2010; Kenny et al., 2019). However, visual fidelity still significantly influences some perceptual outcomes, such as object size perception, which avatar realism can notably impact, especially in highly realistic scenarios (Ogawa et al., 2018).

Recent evidence by Lee et al. (2025) qualifies this view. In a between-subjects VR study, photoreal self-avatars only improved presence when their facial motion was perfectly tracked; any facial-motion lag erased the ownership benefit and re-introduced uncanny discomfort. The finding reinforces the present study's focus on dynamic fidelity rather than static polygon count.

From a perception standpoint, two factors are especially relevant: (i) the *magnitude* of change between LODs (e.g., loss of facial micro-expression rigging), and (ii) the *temporal profile* of the transition (abrupt pops versus gradual morphs). In this experiment, all transitions are abrupt LOD pops by design. Abrupt onsets and motion onsets reliably capture attention and are prioritised by bottom-up salience mechanisms, especially when a single onset is present (Yantis and Jonides, 1984; Abrams and Christ, 2003; Itti and Koch, 2001). No gradual morphing is used so that detectability thresholds reflect unmasked transitions under cinematic conditions. The focus is on non-interactive, non-VR, cinematic presentation to isolate visual detectability without confounds from interaction.

Detectability is not solely a function of stimulus level. Peripheral performance declines with eccentricity, primarily due to crowding rather than acuity, with critical spacing scaling at roughly half the retinal eccentricity (Bouma's law) (Strasburger et al., 2011; Rosenholtz, 2016). Independently of eccentricity, abrupt onsets and motion onsets capture attention and are prioritised by bottom-up salience mechanisms, especially when a single onset is present (Yantis and Jonides, 1984; Abrams and Christ, 2003; Itti and Koch, 2001).

1.3 Why ownership and similarity-based factors might amplify sensitivity

While most existing work focuses on static avatars or virtual reality applications, relatively little about user sensitivity to LOD changes in non-interactive, non-VR contexts such as cinematic animations is known. Additionally, few studies examine how perceptual factors like avatar similarity or the illusion of embodiment (the multisensory feeling that a body is one's own (Carruthers, 2008)) might interact with technical artefacts like LOD transitions. For example, research by Lugrin et al. (2015) demonstrates that avatars do not require a realistic human appearance to elicit a strong sense of virtual body ownership. In some cases, highly human-like avatars may even reduce the illusion of embodiment due to uncanny valley effects when they fail to match users' expectations of their appearance. On the other hand, Slater et al. (2010) showed that the mere adoption of a first-person perspective can significantly enhance the illusion of embodiment, independent of detailed "visuotactile synchronisation". Together, these findings motivate testing whether personalisation and viewing context shift sensitivity to artefacts. Specifically, whether ownership cues lower the detection threshold for fidelity losses.

1.4 Identified Research Gap

Although there is significant research into avatar realism and animation fidelity (Lee et al., 2025; Yamac et al., 2024), there remains a limited understanding regarding how dynamic changes in LOD affect user perception in realistic avatar animations. Previous studies have mainly focused on static scenarios or isolated perceptual assessment within VR or simplified contexts (Kenny et al., 2019; Hodgins et al., 2010). Also, little is known about user sensitivity to real-time fidelity transitions in avatars, especially in non-interactive, cinematic conditions that mirror practical real-world applications.

Moreover, few studies explore how personalisation (i.e. avatars customised to resemble the viewer) impacts perceptual sensitivity towards fidelity changes. Existing evidence indicates personalised avatars strengthen the sense of embodiment and ownership (Waltemate et al., 2018), potentially amplifying sensitivity to visual anomalies. However, how these perceptual and psychological factors interact with dynamic visual fidelity adjustments has not yet been systematically investigated.

This study addresses this research gap by evaluating real-time detection thresholds for fidelity transitions, specifically comparing generic avatars (NPCs) against personalised MetaHuman avatars that resemble viewers.

This taxonomy motivates the methodological choices developed in chapter 5 and is revisited in the contributions (section 8.2).

Table 1.1: Positioning the present study against prior work

Study	Focus	Methodology	Unresolved Gap
Lee et al. (2025)	Immersive vs. non-immersive avatar realism	Static evaluation; VR vs 2-D; personal avatars; questionnaires	No real-time LOD changes; did not test on-the-fly detection
Yamac et al. (2024)	Physicality errors in visual & kinematic cues	Animated sequences; effort/weight ratings; VR	No dynamic LOD transitions; no personalisation aspect
Kenny et al. (2017, 2019)	Animation (re-)targeting inconsistency	Static stimuli; perceptual judgements; weight inference	No real-time LOD change; limited avatar personalisation
Hodgins et al. (2010)	Saliency of motion anomalies	Pre-rendered clips; anomaly ratings	No live dynamic LOD; no ownership manipulation
Benda and Ragan (2021)	Avatar visibility and spatial accuracy	Interactive VR pointing task; observer ratings	No fidelity manipulation; no user-avatar identification
Kang et al. (2022)	Facial LOD and social presence in AR	Remote-collaboration task; low/mid/high face meshes	Only facial region; static LOD; no detection timing
Lugrin et al. (2015)	Anthropomorphism & embodiment	IVBO in VR with stylised vs realistic bodies	No LOD manipulation; ownership studied but not fidelity change
Waltemate et al. (2018)	Personalised scans & body ownership	Photogrammetry self-avatars; VR questionnaires	Strong ownership effect shown, yet no LOD change tested
Nguyen et al. (2024)	Mesh simplification thresholds in VR	Log-linear MOS for vertex count X distance	Distance-based Q only; no self-avatars; no change-detection task
Sun et al. (2025)	Gaussian-splat crowds LOD	Artefact-detection study; motion complexity X screen space	Crowd context; static LOD; not self-resembling avatars
Ogawa et al. (2018)	Hand realism & object-size perception	Hand-scale manipulation in VR reach tasks	Body-part scaling, not LOD; no dynamic transitions
Proposed study	**Dynamic real-time LOD detection with personal avatars**	**Live MetaHuman animation; personalised vs neutral avatars; click-only detection task**	**First to combine dynamic LOD changes with ownership manipulation and real-time detection thresholds**

1.5 Research Questions and Objectives

This study addresses these gaps by examining how easily users detect LOD changes in animated MetaHuman avatars within a non-VR cinematic context and whether sensitivity is elevated when the avatar is personalised. Specifically, the study asks:

- **RQ1 (Detection Thresholds).** How quickly and reliably do users detect real-time GeoLOD changes of animated avatars?
- **RQ2 (Ownership Influence).** Does the self-resemblance of avatars enhance user sensitivity to fidelity changes compared to generic avatars?

By exploring these effects, this research aims to provide practical guidelines for developers seeking to optimise performance and user experience in real-time avatar applications.

To address these questions, the following objectives are proposed:

- **Develop a framework** that dynamically modifies GeoLOD in MetaHuman avatars in real-time, simulating realistic cinematic conditions.
- **Implement a detection task**, where users indicate (via mouse clicks) the exact moment they notice changes in visual fidelity (GeoLOD).
- **Compare detection rates** between self-resembling and generic avatars to quantify the effect of avatar personalisation on perceptual sensitivity.
- **Generate practical guidelines** for developers on the optimal use of personalised avatars and fidelity management strategies in real-time cinematic and gaming environments.

1.6 Thesis roadmap

- **Chapter 2** reviews visual realism, the uncanny valley, and fidelity trade-offs, then summarises similarity/ownership and perspective effects, drawing takeaways for LOD design.
- **Chapter 3** examines how motion governs perception, outlining motion-dominant cues and the interaction between motion and appearance.
- **Chapter 4** surveys perceptual LOD thresholds and resource budgets (geometry and rendering), highlighting distance/screen-space/motion moderators and implications for budget setting.
- **Chapter 5** details the methodology: participants, apparatus and stimuli, viewing conditions, timing, procedure, logging/QC, and the analysis plan.
- **Chapter 6** reports results: sample/data quality, descriptives (per-LOD detection with CIs), psychometric fits, diagnostics/robustness, and an interim summary.
- **Chapter 7** discusses interpretations, relates findings to prior work, translates them into production guidance (practical implications and checklist), and covers limitations and future work.
- **Chapter 8** concludes with key insights and contributions, and final remarks.
- **Appendices** include ethics approval and the consent form.

Chapter 2

What makes an avatar look 'right'?

2.1 Visual realism, the uncanny valley and fidelity trade-offs

The influence of visual realism on how users perceive digital avatars remains a point of contention in VR and computer graphics literature. Usually, high-fidelity rendering heightens immersion and strengthens the sense of presence (Cummings and Bailenson, 2016). However, as Mori et al. (2012) argued, extreme lifelike but imperfect characters risk plunging into an "uncanny valley" that elicits revulsion rather than empathy.

One notable example comes from Lugrin et al. (2015), whose findings suggest that avatars with simplified or stylised appearances, ranging from cartoonish to mechanical, can provoke a comparable, if not stronger, illusion of virtual body ownership (IVBO) than their highly realistic counterparts. Their study observed that participants often felt more embodiment when using less human-like avatars, reporting slightly stronger IVBO experiences than with photorealistic models. Interestingly, the more detailed avatars sometimes caused perceptual confusion, with users describing sensations akin to having "two bodies". This dissonance appears to be related to the Uncanny Valley effects, where minor inconsistencies, such as odd proportions or unnatural clothing, disrupt the illusion of embodiment.

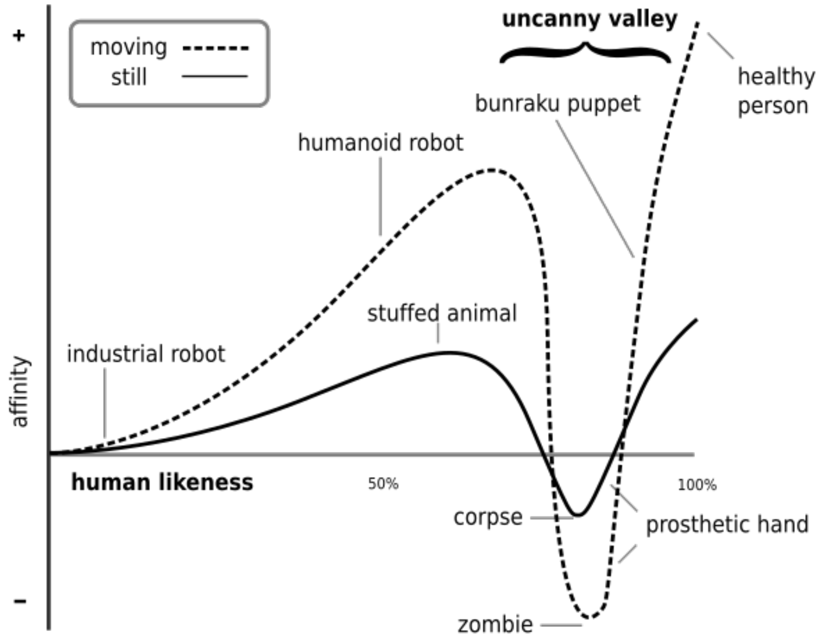


Figure 2.1: The "uncanny valley" as a conceptual relation between human likeness and affect (schematic). Source: *Mori Uncanny Valley.svg* by Smurrayinchester, Wikimedia Commons, licensed CC BY–SA 3.0. Based on Mori (1970/2012). No changes made.

Hodgins et al. (2010) further support this sensitivity to imperfection. Their research shows that even subtle errors in high-fidelity avatars can be more off-putting than significant flaws in stylised ones. Across a series of perceptual tests, viewers consistently rated facial animation errors as particularly distracting despite otherwise fluid full-body motion, confirming that facial anomalies dominate salience and can drag an otherwise polished character into the valley.

Furthering this discussion, Kang et al. (2022) explored how different levels of avatar face detail affect social presence and interpersonal communication in augmented reality (AR). They found that moderate levels of facial detail (mid-LOD) maintained a social presence as effectively as high-detail faces, significantly reducing computational costs. Conversely, excessively simplified avatars (low-LOD) negatively impacted users' emotional understanding, demonstrating the necessity of maintaining essential facial cues even at reduced fidelity.

Together, these studies underline that visual realism is beneficial only when the fidelity of facial cues matches users' perceptual expectations.

2.2 Similarity, ownership and the role of perspective

Research consistently demonstrates that avatar appearance and the level of embodiment significantly influence user perceptions and experiences within virtual environments. Slater et al. (2010) showed that adopting a first-person perspective (1PP) was sufficient to generate a strong sense of body ownership, even without synchronous visuotactile stimulation. Maselli and Slater (2013) support this by stating that a 1PP over a co-located humanoid body is necessary for the full-body ownership illusion.

Contrary to expectations, highly realistic avatars do not always improve the sense of ownership. For example, Lugrin et al. (2015) found that less realistic avatars, such as simplified or stylised representations, can produce stronger feelings of embodiment than photorealistic avatars. This effect is partly explained by the Uncanny Valley, where minor inconsistencies in highly realistic avatars (e.g., unnatural proportions or movements) create discomfort and cognitive dissonance, disrupting the sense of ownership (Lugrin et al., 2015).

Personalisation adds another layer. Waltemate et al. (2018) demonstrated that avatars generated from fast photogrammetry scans and matched to the participants' own faces significantly increased body ownership scores, presence, and perceived dominance compared to generic counterparts.

This implies that self-similarity can offset the risk of triggering the Uncanny Valley. However, the match must be precise, as near-misses are worse than stylisation.

Table 2.1: Evidence that precise self-matches outperform near-miss realism.

Study	Conditions	Metric	Finding
(Waltemate et al., 2018)	Generic-model vs. Generic-scan vs. **Personal-scan**	IVBO Acceptance	Means 2.66 –> 3.11 –> 4.02; personalised > both generics ($p < .01$).
(Lugrin et al., 2015)	Human vs. Robot vs. Box	"Two bodies" item	Human avatar significantly higher eeriness than both stylised ($p < .05$).
(Mori et al., 2012)	Conceptual curve	Affinity vs. likeness	Figure 1: affinity drops for almost-human with defects (uncanny valley).
Hodgins et al. (2010)	Realistic face vs. stylised body errors	Saliency ratings	Micro facial glitches judged most disturbing despite flawless body motion.

2.3 Takeaways for LOD design

Visual realism, while often seen as a technical achievement, does not guarantee better outcomes. Its impact hinges on how well visual appearance aligns with motion behaviour and user expectations. When there is a mismatch (i.e. when a realistic avatar moves unnaturally or fails to match a user's internal model), presence, ownership, and embodiment can suffer. Consequently, LOD systems must balance perceptual rather than purely geometric priorities. Coarse or stylised models may be enough (or even better) for non-personal avatars, whereas personalised scans need higher budgets because users notice small discrepancies.

Chapter 3

How motion governs perception

3.1 Motion-dominant cue findings

How virtual characters move plays a critical role in how they are perceived. Across research in psychology, animation, and VR, motion has consistently been shown to guide perception, even overriding visual realism when discrepancies arise (Kenny et al., 2019; Yamac et al., 2024; Ennis et al., 2010).For example, Kenny et al. (2019) found that participants based judgements about actions, like estimating object weight or throw distance, more on movement patterns than on how the characters appeared, even when body shape and motion did not align due to retargeting.

Yamac et al. (2024) similarly showed that when motion and visual features clashed, users leaned heavily on kinematic cues to assess physicality, with perceived realism dropping when expected dynamics were violated. Supporting this, Ennis et al. (2010) observed that in multisensory settings, body motion outweighed both sound and facial cues in conveying communicative intent.

Kenny et al. (2019) ran three experiments to explore how mismatched motion and body shape influence perception. Even with light and heavy body types swapped with incongruent motion data, participants did not flag the animations as off, nor did ratings of eeriness or humanness change. However, when asked to judge the results of those actions (i.e. the apparent weight of an object), participants responded primarily to the motion data. This inability to consciously detect inconsistency does not imply perceptual indifference. The same study showed that although participants failed to identify mismatched stimuli in Experiments 1 and 2, they still displayed systematic differences in judgments about object weight and throwing distance in Experiment 3. These findings suggest that our visual system tends to resolve inconsistencies by favouring coherent motion-based interpretations over visual congruence (Kenny et al., 2019). When a lightweight motion drove a heavyweight body, participants judged the object being lifted as heavier, and vice-versa. This "cognitive absorption" supports the idea that the visual system preserves global coherence by attributing mismatches to external forces rather than flagging the animation as faulty. Any LOD transition that disturbs motion-shape congruence therefore risks altering perceived action outcomes even when the error is not consciously detected.

This aligns with Runeson and Frykholm (1983), who argue that kinematic patterns specify underlying forces and intentions in a lawful, perceivable way. While their original claim supported direct perception, Kenny et al. (2017, 2019) suggest that observers may be attuned to these relationships through lived experiences and exposure.

3.2 When motion meets appearance

Yamac et al. (2024) showed that deviations from expected motion (i.e., reduced joint velocity or unnatural acceleration) significantly undermined perceived realism, even when avatars were rendered with high visual fidelity. These distortions in movement were reliably detected across realistic and stylised characters, highlighting the visual system's sensitivity to kinematic anomalies and reinforcing motion as a dominant perceptual cue (Yamac et al., 2024).

Ogawa et al. (2018) provided complementary insights into how avatar realism interacts with motion to affect perception. They demonstrated that avatar hand realism and size can significantly influence perceived object size. When avatar hands were realistically rendered and enlarged, participants perceived grasped objects as noticeably smaller, illustrating how subtle motion-related interactions with avatars significantly modulate perceptual outcomes. Notably, the influence of motion-related body scaling was significantly weaker with less realistic avatars, indicating that realism amplifies the perceptual impact of motion cues.

Additionally, Kang et al. (2022) demonstrated the critical role facial detail plays in emotional understanding during interactions. They showed that while moderate levels of facial detail (mid-LOD) effectively maintained emotional comprehension, low-detail faces significantly hindered emotional decoding. This indicates that although motion cues are important, detailed visual features, particularly facial expressions, are essential for effective emotional and communicative exchanges in collaborative virtual settings.

Sun et al. (2025) investigated this perceptual sensitivity further in the context of crowd avatars rendered using Gaussian splatting. They found that the number of splats had the strongest effect on artefact detection, with coarse avatars (3k Gaussians) producing a visibly degraded appearance. Also, when avatars were rendered at 40% of their original 406-pixel height, detection accuracy dropped, suggesting that lower screen resolution can mask low-fidelity artefacts.

Taken together, these studies indicate that when an avatar's movement is plausible, even aggressive geometric simplification can go unnoticed, whereas complex kinematics, especially with high-resolution surfaces, make fidelity loss noticeable.

Crucially, observers can fail to notice a mismatch while still reacting to it behaviourally. Kenny et al. (2019) found that participants rarely declared incongruent shape-motion pairings "unnatural" (Experiment 1-2), yet their weight and effort judgements shifted systematically in line with the underlying motion (Experiment 3). Hodgins et al. (2010) also showed subtle-eye gazing and blink errors in high-fidelity faces produced strong discomfort ratings even when participants could not pinpoint the defect. Subtle anomalies in realistic avatars attracted heightened attention and discomfort, even when they did not raise the level of conscious detection. Participants were most sensitive to facial irregularities, suggesting a hierarchy in perceptual salience that prioritises social and emotional cues.

Table 3.1: Empirical evidence that motion salience and surface realism modulate detectability of LOD shortcuts.

Study	**Motion / kinematic manipulation**	**Surface-fidelity context**	**Key finding**
Kenny et al. (2019)	Light-vs-heavy body shapes paired with incongruent motion (Exp 1-2)	Photoreal body meshes, no facial detail	Incongruence rarely noticed; eeriness unchanged, showing smooth motion masks shape mismatch.
Yamac et al. (2024)	Reduced joint velocity / unnatural acceleration	High-fidelity body model	Kinematic errors sharply cut realism ratings even when visuals were flawless, so salient motion reveals flaws.
(Ogawa et al., 2018)	Reaching with enlarged vs normal hand size	Stylised versus realistic hand render	Body-scaling effect on perceived object size nearly doubled with realistic hand, proving surface detail amplifies motion cues.
(Sun et al., 2025)	Complex martial-arts motion vs cyclic jogging	Gaussian-splat avatars at 202 k vs 3 k splats; 100 % vs 40 % pixel height	Complex motion increases artefact detection (3 k splats); distance (40 %) masks same artefacts-motion salience and resolution interact.
(Kang et al., 2022)	Normal conversational gestures (held constant)	Mid-LOD (4 257 verts) vs Low-LOD (252 verts) faces	Mid-LOD faces maintain emotional comprehension; Low-LOD faces lose it despite identical motion, showing low facial detail can expose limits when cues are socially salient.

3.3 Summary

Motion is the main cue people use to judge virtual characters. When motion looks natural, users barely notice even heavy cuts in geometry. However, if the motion feels off, or detailed textures expose tiny flaws, those cuts jump out. LOD systems must consider how noticeable motion is, not just how detailed the model is, before trimming detail.

Chapter 4

Perceptual LOD thresholds and resource budgets

4.1 Geometry-based thresholds (mesh simplification)

Nguyen et al. (2024) provided quantitative insights into perceptual thresholds for VR LOD adjustment. Their study demonstrated that how users' perceived quality is highly dependent on viewing distance and geometric complexity. For highly detailed meshes (>100k vertices), simplifications of up to 95% (i.e 100k $\rightarrow$ 5k vertices) remain perceptible but not annoying (MOS $\geq$ 4) at medium-to-far distances ($\approx$10-20 m). The authors utilised a linear model with log-transformed predictors that predicts subjective quality from the natural logarithms of vertex count (Q) and distance (D), plus an interaction term, offering an operational rule for real-time systems (Equation 4.1).

$$\text{MOS} = \alpha \times \ln(Q) + \beta \times \ln(D) + \gamma \times \ln(Q) \times \ln(D) + \delta \tag{4.1}$$

These findings highlight the practical value of dynamic LOD algorithms that can aggressively cull geometry without degrading user experience, provided simplification respects perceptually relevant parameters such as distance.

4.2 Rendering-based thresholds (Gaussian avatars & facial detail)

Recent advancements in Gaussian-based crowd rendering by Sun et al. (2025) emphasise the nuanced relationship between visual detail and perceived quality. Reducing the number of Gaussians from 202k (gold standard) to 3k significantly increased artefact detectability, and complex martial-arts motion yielded higher detection accuracy than cyclical jogging. Distance alone showed no main effect, yet did interact with LOD: at 40% pixel height (around 7.5 m), the lowest-detail avatars became markedly harder to discriminate, indicating that screen-space resolution can mask errors at extreme distances.

Complementary evidence comes from Kang et al. (2022), who explored avatar-face LOD in Augmented Reality (AR) collaboration. Moderate facial detail (4,257 vertices) preserved social presence and interpersonal attraction as effectively as high-detail faces, whereas extreme simplification to 252 vertices produced significant declines in emotional understanding and trust, especially in socio-emotional tasks. These results collectively underscore that acceptable fidelity depends not only on distance but also on the communicative salience of the rendered region.

4.3 Context moderators: distance, motion complexity and socio-emotional salience

Across both mesh simplification and Gaussian rendering experiments, the detectability of LOD reductions is not fixed but shaped by task context. Nguyen et al. (2024) show a steady increase in MOS with greater viewing distance. However, the size of this gain depends on the starting complexity of the model. Low-vertex meshes continue to look degraded at close range, whereas highly detailed meshes quickly become "good enough" beyond ($\sim$10)m. Sun et al. (2025) likewise show that increasing distance suppresses artefact visibility, but only for the most aggressively simplified (3k Gaussian) avatars; at higher LODs, distance exerts little protective effect. Crucially, their data reveal that complex, non-cyclic motions intensify observers' sensitivity to LOD gaps.

A complementary mechanism comes from *observer-induced* motion. In an immersive VR 2IFC study, Petrescu et al. (2023) asked participants to make yaw head rotations while judging which interval contained a lower-quality model. LOD degradation (quadric-error simplification) was applied only when rotational head velocity exceeded a threshold within the interval. Psychometric fits (75% correct) showed that users tolerated approximately a four-fold polygon reduction even under the "slow" rotation regime, and tolerated *more* degradation at higher rotational velocities (fast > slow; paired $t = 2.71$, $p = 0.008$), consistent with speed-dependent masking from retinal motion. This supports the view that motion can *reduce* detectability when the motion increases retinal image velocity (e.g., fast head movements), even when observers are actively looking for quality loss. **Scope note.** This effect concerns retinal-motion masking from self/camera motion rather than the *kinematic complexity* of an articulated target, and was measured in VR rather than cinematic desktop viewing. It should therefore be treated as convergent but context-bound evidence.

Finally, Kang et al. (2022) demonstrate that when an AR task emphasises emotional exchange, coarse facial meshes (252 vertices) sharply reduce perceived affective understanding and interpersonal attraction, whereas goal-oriented tasks tolerate the same simplification.

Collectively, the evidence supports a conditional tolerance rule. Aggressive LOD cuts are acceptable when the object is far away, its movement is simple or peripheral, and the region carries low socio-emotional weight. Violating any one of these conditions (i.e. close inspection, intricate kinematics, or face-to-face interaction) renders simplification immediately noticeable.

4.4 Implications for LOD budgets

Quantitative guidance for budget setting comes from two key sources:

- **Predictive tools**
  - MOS model of Nguyen et al. (2024): maps vertex count and viewing distance to expected quality, yielding explicit vertex-count targets.
  - LOD-pixels interaction matrix of Sun et al. (2025): relates Gaussian count and on-screen extent to artefact detectability, defining Gaussian budgets for crowd avatars.
- **Tiered allocation strategy**
  - *Maximal detail* – facial regions of primary or self-relevant avatars (Kang et al., 2022).
  - *Intermediate detail* – secondary interactive characters (Kang et al., 2022).
  - *Aggressive reduction* – distant background assets (Nguyen et al., 2024; Sun et al., 2025).
- **Transferability**
  - The same numeric thresholds provide a first-order benchmark for cinematic, non-VR scenes where the camera is fixed but motion may remain complex (Sun et al., 2025; Nguyen et al., 2024).

Chapter 5

Methodology

5.1 Overview and rationale

This chapter details the experimental method used to estimate perceptual detection thresholds for GeoLOD *changes* in animated MetaHuman avatars under non-interactive, non-VR, cinematic presentation. A fixed-reference Method of Constant Stimuli was implemented as an *alternating* sequence of reference and test segments ($\ldots 0 \rightarrow X \rightarrow 0 \rightarrow Y \rightarrow \ldots$) (Gescheider, 2013; Kingdom and Prins, 2016). Participants indicate the moment they detect a change with a mouse click. The primary outcome is the LOD level at which detection becomes more likely than not (the 50% threshold), estimated by fitting logistic psychometric functions to trial-level responses. Where applicable, thresholds are compared between a Customised and a neutral avatar to test whether ownership cues shift sensitivity (Lugrin et al., 2015; Slater et al., 2010; Lee et al., 2025). The section that follows specify participants, apparatus, procedure, logging, data-quality criteria and analysis in full.

5.2 Participants

Ten volunteers (university students), 6 women and 4 men, took part in the study. Participants offered written informed consent for the experiment. Recruitment used convenience sampling via personal contacts within the university. No mailing lists or participant pool were used. Participation was uncompensated. Inclusion criteria were age $\geq$ 18, normal or corrected-to-normal vision, and fluency in English. Individuals who had taken part in pilot testing were excluded. No further exclusions were necessary. No participants requested accessibility adjustments, and none were required. For the customised condition, each participant created a Customised avatar using MetaHuman Creator prior to the main task. Participants were free to adjust facial features and other appearance parameters to enhance perceived similarity and ownership. The neutral avatar (generic identity) was used as the comparison condition. Each session lasted approximately 50 minutes, including a short 10-minute break at the midpoint (i.e. after completing the neutral-avatar block and before the customised-avatar block, or vice-versa). Written informed consent was obtained prior to participation, and data were recorded under a pseudonym participant ID. Full demographics are reported in Table 5.1. Ethical approval for the study was granted under WMG's Supervisor Delegated Approval (SDA) procedure at the University of Warwick (see Appendix .1 for the approval notice).

Table 5.1: Participant demographics and background. Values are counts unless otherwise stated.

Measure	**Summary**
Sample size	10 participants; all completed both avatar blocks
Age (years)	$M = 25.3$, $SD = 2.91$, range 22–32
Gender identity†	Women 6; Men 4; Non-binary/Other 0; Prefer not to say 0
Vision	Normal 4; Corrected-to-normal with glasses 6; Contacts 0
Weekly gameplay	Median 2.5 h (IQR 0–10.25); range 0–14
3D/engine experience	None 5; Student/enthusiast 3; Professional 2
Unreal Engine exposure	Yes 5; No 5
MetaHuman exposure	Yes 6; No 4

† Gender categories were self-reported; Counts are reported as provided by participants.

5.3 Design

The study used a within-subjects design so that each participant served as their own control across conditions. The primary manipulation was GeoLOD during a brief Test interval, with levels $X \in \{1, \ldots, 7\}$.

For the face mesh, triangles decrease from 64,094 (LOD 0) through 6,872 (LOD 3; -89.3%) to 356 (LOD 7; -99.4%). For the body mesh, from 28,900 (LOD 0) to 890 (LOD 3; -96.9%). Full counts and per-level percentages are summarised in Table 5.2.

Percentages differ across meshes because reductions are computed relative to each mesh's own LOD 0, and because MetaHuman employs separate LOD schedules for face (0-7) and body (0-3). With head-and-shoulders framing, face changes are therefore expected to dominate detectability.

LOD 0 acted as a constant high-fidelity reference and was not entered into psychometric fitting. Presentation used a fixed-reference, *alternating* sequence of reference and test segments $(\ldots 0 \rightarrow X \rightarrow 0 \rightarrow Y \rightarrow \ldots)$ with abrupt LOD pops at segment boundaries. The primary dependent variable was a binary detection response (mouse click) during the Test interval. Reaction time (RT) from test onset to first click was recorded as a secondary outcome. Two experimental factors were varied within participants:

- **Avatar personalisation:** Customised vs Neutral avatar
- **Animation state:** Idle vs Walking clips of equal duration and viewpoint.

Conditions were organised in blocks. By default, animation state was nested within avatar blocks (e.g. Custom-Idle, Custom-Walking, break, Neutral-Idle, Neutral-Walking), but the ordering of avatar blocks was counterbalanced across participants (AB/BA) to control for learning and fatigue. Within each block, GeoLOD levels were randomised across trials. No accuracy feedback was given.

5.4 Hypotheses

The following research questions are addressed:

Hypotheses:

- **H1** Detection probability increases monotonically with GeoLOD (more degraded levels are easier to detect), yielding a measurable 50% detection threshold.
- **H2** Personalised/self-resembling avatars produce lower detection thresholds (greater sensitivity) than neutral avatars, consistent with ownership effects on perception.

Formal hypotheses are tested in the Analysis Plan (section 5.9).

5.5 Apparatus and Stimuli

5.5.1 Software

Stimuli were rendered in **Unreal Engine 5.6.0** (Games, 2025b) using MetaHuman assets (Games, 2025a). GeoLOD was controlled via `ULODSyncComponent::ForcedLOD` at segment boundaries (Ref $\rightarrow$ Test $\rightarrow$ Ref). The experiment runner and logger were implemented in C++ within the same project build.

5.5.2 Hardware and display

The experiment was conducted on a laptop PC (**CPU: Intel(R) Core(TM) Ultra 7 155H**, **GPU: NVIDIA GeForce RTX 4050/6GB VRAM**, **RAM: 32GB**, **OS: Windows 11**). Visuals were presented on a **15-inch** monitor at **1920×1200** and **120Hz** in exclusive full-screen mode. Participants responded with a standard USB mouse; audio was disabled.

5.5.3 Viewing conditions

Participants were seated at approximately **20 inches** from the display in a dimly lit, quiet room. The display height was adjusted so that the avatar's body was roughly centred on screen (central framing reduces peripheral-vision crowding confounds, which degrade change detection away from the fovea (Strasburger et al., 2011; Rosenholtz, 2016), thereby ensuring that transitions occur in high-acuity vision rather than being masked in the periphery.

To standardise retinal location, the avatar was centred on screen. A fixed camera and viewing distance were used. Switches were triggered at a consistent point in the motion cycle to avoid phases with unusually large silhouette change.

Figure 5.1: Experimental setup.

5.5.4 Avatars

Two avatar identities were used:

- **Neutral:** a generic MetaHuman with standard clothing, held constant across all trials and LODs.
- **Customised:** created by each participant in MetaHuman Creator prior to the task, adjusting facial morphology, clothes, and other appearance parameters to increase perceived similarity and ownership.

Figure 5.2: Example of avatar stimuli used in the experiment. Neutral MetaHuman (right) and a representative Customised MetaHuman (left), shown under identical lighting and camera framing.

(a) LOD 0 (b) LOD 1 (c) LOD 2 (d) LOD 3

(e) LOD 4 (f) LOD 5 (g) LOD 6 (h) LOD 7

Figure 5.3: Visual progression across MetaHuman geometry LODs (engine LOD 0-7) under identical camera and lighting. Each panel shows the same frame from a neutral idle. Detailed reduction in triangle counts in Table 5.2

5.5.5 Animations

Each avatar performed two motion states: **Idle** and **Walking**. Animations were authored as Unreal Level Sequences and played continuously. The onset of each GeoLOD segment was timed by the controller. There was no audio. The same clips and timings were used for both Neutral and Customised avatars.

5.5.6 LOD control

GeoLOD was manipulated by setting `ULODSyncComponent::ForcedLOD` at the start of each segment. **LOD 0** served as the high-fidelity reference and **LOD 1-7** as test levels. All changes were **abrupt pops** at segment boundaries. Where necessary, meshes sharing facial and body detail were grouped to receive the same LOD index. Sub-systems not governed by LODSync (e.g., certain grooms) were held constant to avoid confounds. Two LOD 0 segment types served distinct roles. Pre-test LOD 0 segments provided a conventional **false-alarm (FA)** estimate, whereas the **return LOD 0** segments immediately after a detected test served **return-validation** only. Both LOD 0 types were excluded from psychometric fitting.

By convention here, larger GeoLOD indices denote more aggressive simplification (lower geometric detail). LOD 0 is the highest-fidelity reference. For the body mesh, triangle counts decreased from 28,900 (LOD 0) to 890 (LOD 3), a 96.9% reduction. For the face mesh, counts decreased from 64,094 (LOD 0) to 356 (LOD 7), a 99.4% reduction. Table 5.2 summarises the counts and percentage reductions used in the stimuli.

Figure 5.4 illustrates the general runtime logic by which real-time engines assign LODs as a function of distance or projected screen size. Although this experiment overrides this process by forcing specific GeoLODs at segment boundaries, the diagram clarifies the typical control pathway that motivates such manipulations.

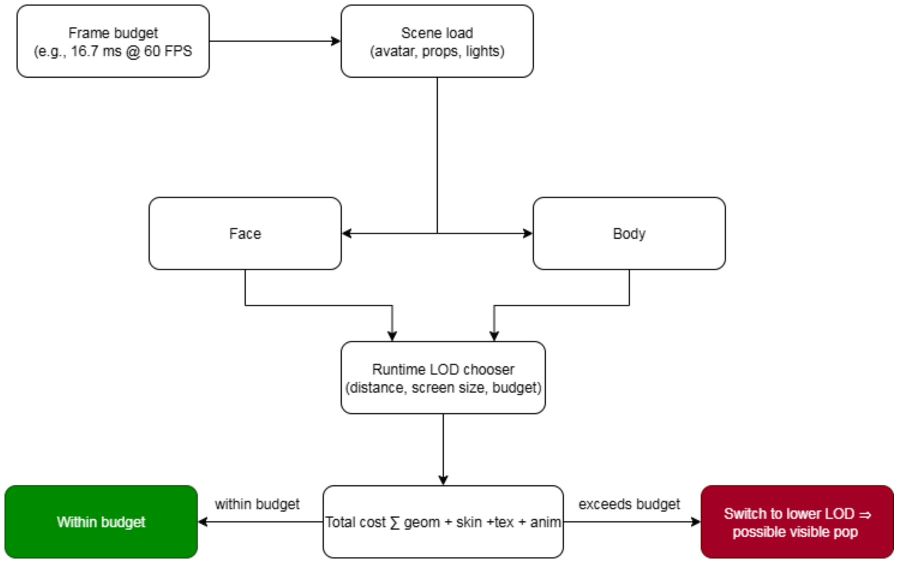


Figure 5.4: Illustrative LOD-by-distance rule. As on-screen size shrinks with distance, the engine steps down geometry/texture LODs to meet the frame budget. Abrupt steps risk perceptible pops.

5.5.7 Visual context and camera

A static camera framed the head and upper torso against a neutral background. Camera position, field of view and lighting were fixed across all conditions and LODs. Post-processing was minimal and constant (no exposure or colour grading changes).

5.5.8 Timing and trial generation (summary)

The experiment presented an **alternating sequence of reference and test segments** rather than packaging each trial as a Ref→Test→Ref triplet. Specifically, a shuffled list of test GeoLODs $\{1, \ldots, 7\}$ (repeated k times) was generated. A reference segment (LOD 0) was inserted before each test, and a final LOD 0 concluded the run. Each segment's dwell time was sampled uniformly in the range **3-7 s**. **FA estimates** used only the pre-test LOD 0 segments. Return LOD 0 segments were not counted as FA, they were used solely to validate detections. Test GeoLOD levels were randomised within blocks. The random order was held only for the duration of a session.

Table 5.2: Triangle counts by LOD for body and face meshes (percent reduction relative to LOD 0).

	Face mesh (triangles; reduction %)							
LOD	0	1	2	3	4	5	6	7
Triangles	64,094	34,514	17,274	6,872	3,504	1,524	710	356
Reduction	0%	46.2%	73.0%	89.3%	94.5%	97.6%	98.9%	99.4%
	Body mesh (triangles; reduction %)							
LOD	0	1	2	3				
Triangles	28,900	7,353	3,105	890				
Reduction	0%	74.6%	89.3%	96.9%				

5.6 Procedure

5.6.1 Session Flow

Participants attended a single lab session. After informed consent and a short demographics questionnaire, they received instructions and completed a brief warm-up. The main experiment comprised two blocks (avatar personalisation: *Neutral*, *Customised*), each containing two motion states (*Idle*, *Walking*) presented as separate sub-blocks. A short planned break separated the avatar blocks. Finally, participants were debriefed and thanked for their time. Total session time was approximately **50 minutes**, including a **10-minute break** at the midpoint.

5.6.2 Instructions

> *Press the mouse as soon as you notice any change in the avatar's visual detail. Do not guess; if you are unsure, withhold your response. Keep your gaze on the avatar.*

Instructions were read in British English. No performance feedback was provided during the task.

5.6.3 Warm-up (excluded from analysis)

To familiarise participants with timing, response mapping, and the types of geometric degradation. The warm-up comprised **five** coarse GeoLOD examples in fixed order: $0, 2, 4, 6, 7$. Warm-up trials sampled both avatar conditions (Neutral and Customised) and used the same abrupt LOD pops and response mapping as the main task. These segments matched the main task in presentation but were *not* analysed.

5.6.4 Trial structure

Presentation alternated between a **reference** segment $(\mathrm{LOD}\,0)$ and a **test** segment $(\mathrm{LOD}\,X,\ X \in \{1, \ldots, 7\})$, followed by the next reference, and so on. Segment durations were uniformly sampled between **3-7 s**. Participants were instructed to click *as soon as* they detected a change. The first click within a segment was recorded, and additional clicks were ignored until the next segment. A minimum RT filter was applied during analysis to remove anticipatory responses. LOD 0 segments were excluded from psychometric fitting. Only pre-test (LOD 0) segments contributed to FA estimates. For each avatar *and* motion state, each GeoLOD level was repeated **k** times (planned $k = \mathbf{8}$), yielding $7 \times k$ test segments interleaved with reference segments. The immediate return segment (LOD 0) following each test provided an internal consistency check (return-validation). FA was reported descriptively and not used for participant-level exclusion.

5.6.5 Randomisation and counterbalancing

The order of avatar blocks was **counterbalanced** across participants (Customised→Neutral vs. Neutral→Customised) to control for learning effects and fatigue. Within each block, GeoLOD levels were **randomised** across test segments. Reference segments always preceded tests. No accuracy feedback was given.

5.6.6 On-screen display and feedback

The interface was intentionally minimal. No correctness or timing feedback was given during the experiment. All other presentation parameters (camera, lighting, background) were fixed across trials and conditions, as specified in section 5.5.

5.7 Data logging and quality control

5.7.1 Logged fields (segment-level schema)

Logging occurred at the *segment level*: each row corresponds to the onset and display of a single LOD segment (either the reference LOD 0 or a test GeoLOD $X \in \{1, \ldots, 7\}$). A logical MoCS "trial" (Ref→Test→Ref; $0 \rightarrow X \rightarrow 0$) therefore spans three consecutive rows in the log. The columns present in all exported CSV files are:

Field	Type	Description
`ParticipantID`	string	Pseudonymous code (e.g., P01).
`Trial`	int	Running index of segments within a session.
`GeoLOD`	int	GeoLOD displayed for the segment: 0 (reference) or 1-7 (test).
`DeltaGeoLOD`	int	Logged difference to previous segment. *Not used in analysis*; values are not reliable in MoCS rows.
`DistanceBin`	factor	Viewing-distance bin (fixed at "3m" in this study).
`MotionType`	factor	Motion label from the runner (present but not diagnostic in current logs).
`Phase`	factor	`WarmUp` or `MoCS`. Only `MoCS` is analysed.
`Detected`	int	Binary response within the segment (1 = click occurred; 0 = no click).
`RT_ms`	float	Reaction time (ms) from segment onset to the first click; 0 if no click.
`AvgMS_Before`	float	Mean frame time (ms) in a short window before segment onset.
`AvgMS_After`	float	Mean frame time (ms) in a short window after segment onset.
`TimeStim`	float	Monotonic engine timestamp at segment onset.
`TimeResp`	float	Monotonic engine timestamp at first response (if any).

Table 5.3: Segment-level logging schema found in exported CSVs

Notes. (i) LOD 0 segments reflect the reference presentation. **Pre-test LOD 0 segments** were used to estimate FA, **return LOD 0 segments** were used for **return-validation**. Both were excluded from psychometric fitting. (ii) `DeltaGeoLOD` is not used because it does not reliably encode the previous LOD in MoCS; the analysis reconstructs condition labels directly from `GeoLOD` and block metadata. (iii) `MotionType` is retained for completeness but motion state (Idle/Walking) is defined by the block in which the file was collected.

5.7.2 Primary scoring (per-segment)

Analyses operate on `MoCS` rows only. For test segments ($X \in \{1, \ldots, 7\}$), the binary response `detect` was coded from `Detected` (1 = click; 0 = no click). For pre-test reference LOD 0 segments (the LOD 0 immediately preceding a test), any click counted toward the participant's FA rate. Return LOD 0 segments were not counted as FAs. Instead the return-validation click % was computed as the proportion of detected Test segments (GeoLOD > 0 with a click) whose immediate return (LOD 0) also contained a confirming click. Because presentation alternated $0, X, 0, Y, \ldots$, a *return-validation* rule was applied: if a detected test segment (GeoLOD > 0, `Detected=1`) was immediately followed by a return LOD 0 segment with `Detected=0`, the test detection was flagged as suspect and excluded from modelling (Section 5.8). Only the first click within a segment was considered.

- **Detect (1/0):** coded from `Detected` $(1 = \text{click}; 0 = \text{no click})$ with a minimum-RT filter (below).
- **RT (ms):** `RT_ms` for detected segments; undefined otherwise.

For reference segments (LOD 0) and explicit warm-up rows:

- **False alarms (FA):** any click during LOD 0 segments (`Detected=1`) is counted toward the participant's FA rate.

Trials flagged by the return-validation rule or anticipatory responses were excluded from analysis and not replaced. All models were fitted to the remaining trial-level data without imputation or reweighing; exclusions were applied symmetrically across conditions. Exclusion counts by condition are reported in Section 6.1.

5.7.3 Quality control rules

Before modelling, the following pre-specified, trial-level QC were applied to the segment logs:

1. **Phase / stimulus scope.** Retain *MoCS* rows only; drop *WarmUp*. Exclude all LOD 0 rows from psychometric fitting. Pre-test LOD 0 segments were retained solely to estimate false-alarm (FA) propensity; return LOD 0 segments were used only for return-validation.

2. **Anticipation filter.** Discard clicks with RT $< 150\,\text{ms}$.
3. **Timing integrity.** Exclude rows with missing fields or non-monotonic engine timestamps (`TimeStim`, `TimeResp`).
4. **Performance sanity check (contingency).** Engine frame-time fields were logged to identify potential transients. No trials were removed on this basis in the present dataset.
5. **Return-validation.** For each detected test segment (GeoLOD > 0, `Detected`=1), if the immediately following return LOD 0 segment had `Detected`=0, the detection was flagged as unreliable and the test row was excluded from modelling.
6. **Participant-level exclusions.** None were applied in confirmatory analyses on the basis of FA or miss rate. Misses are expected under MoCS at low LODs, and the "return-validation %" metric is not a false-alarm rate.

5.7.4 Derived analysis dataset

After QC and return-validation filtering, the trial-level dataset for modelling comprises one row per MoCS *test* segment with fields: {`ParticipantID`, `GeoLOD` (1–7), `detect`, `RT_ms`, `block labels (avatar, motion)`}. No aggregation is performed prior to modelling. Observed proportions by LOD are shown only for visualisation.

5.7.5 Planned exclusions in modelling

LOD 0 rows are never entered into the psychometric fit and serve only to estimate false alarms. Logistic models (logit link) are fit to MoCS test segments as specified in section 5.9. Robustness checks focus on fit validity (excluding condition-level mis-fits with non-positive slopes or non-convergence) and confirm that pooled thresholds are stable. FA-based participant exclusions were not performed because FA is defined on pre-test LOD 0 only and return-validation is not a false-alarm measure.

5.7.6 Reproducibility

For each participant and condition (avatar $\times$ motion), raw CSVs, refined analysis CSVs, and diagnostic figures (psychometric plots) are retained in a structured directory per participant. Engine clock times (`TimeStim`/`TimeResp`) are recorded to allow auditing of temporal order; wall-clock time is not required for the within-session analyses.

Beyond archival of per-participant files, the complete analysis pipeline is provided as a single Jupyter notebook (Jardim, 2025). The notebook is self-contained and parameterised with project-relative paths (`RAW_DIR = data/raw`, `OUT_DIR = data/processed`). On execution it regenerates all tables and figures used in the manuscript into `data/processed/tables` and `data/processed/figures` without manual intervention. To ensure deterministic outputs across machines, the notebook fixes the process timezone to UTC and sets explicit seeds for Python's `random` and NumPy. The quality-control and exclusion rules (including return-validation and the slope/threshold plausibility checks) are implemented in code exactly as described in the Methods, so that running the notebook on the raw Excel logs reproduces the reported numbers.

All code and input spreadsheets necessary to regenerate the analyses are available in a public repository at github.com/xpt07/lod-detection-thresholds.

5.8 Data preprocessing

All analyses operated on the exported segment-level logs. Preprocessing proceeded as follows. First, rows from the `WarmUp` phase were removed, and only `MoCS` rows were retained. Second, reference segments at LOD 0 were excluded from psychometric fitting. Only pre-test LOD 0 segments were retained to estimate FA rate. Third, return-validation filter was then applied to remove suspect detections. Specifically, for each `MoCS` row with GeoLOD > 0 and `Detected`=1, if the immediately following row corresponded to a return LOD 0 segment with `Detected`=0, the detected test row was dropped.

No aggregation was performed prior to modelling: each row corresponds to one test segment with fields {`ParticipantID`, `GeoLOD`, `detect`, `RT_ms`, `motion`, `avatar`}. For visualisation only, observed detection proportions and Wilson 95% confidence intervals were computed per LOD and condition.

5.9 Analysis plan

5.9.1 Primary analyses

For each participant and condition (avatar $\times$ motion), a logistic psychometric function with logit link was fitted to *return-validated* trial-level responses:

$$\Pr(\text{detect} = 1) = \text{logit}^{-1}\big(b_0 + b_1 \cdot \text{GeoLOD}\big), \qquad x_{0.5} = -\frac{b_0}{b_1}.$$

Only test segments $X \in \{1, \ldots, 7\}$ were used for fitting. LOD 0 rows were used solely to quantify false-alarm propensity. The *detection threshold* $x_{0.5}$ is the GeoLOD at 50% detection.

Uncertainty and reporting. Confidence intervals for $x_{0.5}$ were computed with the *delta method*, i.e., a linear error-propagation of the model's coefficient covariance to $-b_0/b_1$. In practice, the model supplied a standard error for $x_{0.5}$ with this approximation, and 95% CIs were reported as $\hat{x}_{0.5} \pm 1.96\,\text{SE}(\hat{x}_{0.5})$. Fits with non-positive slope ($b_1 \leq 0$) or failing pre-registered QC (anticipation filter and return-validation) were excluded. Model adequacy was checked with predicted-versus-observed plots. Low detection at small LOD changes is expected for sub-threshold stimuli and is reflected by wider binomial CIs in the per-level descriptives (Fig. 6.2).

5.9.2 Justification of trial method and modelling choices.

The fixed-reference Method of Constant Stimuli (MoCS) was used because it evenly samples across all test LODs, allowing estimation of a complete psychometric function rather than converging only on a single threshold. This approach is the classical method for detecting both sub-threshold misses and near-ceiling performance, which are essential for estimating slope and comparing conditions (Gescheider, 2013; Hautus et al., 2021). A one-interval yes/no detection task was chosen to mirror the real-time perceptual decision of noticing a pop, even though such tasks are more criterion-dependent than two-alternative forced choice (2AFC) designs, which are less biased and more efficient (Hautus et al., 2021). Abrupt pops were used by design, as the study sought to test the perceptual salience of unmasked transitions in their worst-case form, rather than blended changes. For analysis, a logistic function (one of several standard sigmoidal forms) was fitted, via maximum likelihood (Wichmann and Hill, 2001). Thresholds were defined as the 50% detection point, the conventional measure in yes/no designs, while noting that alternative criteria (e.g., 75% correct in 2AFC, or d'-based thresholds) are also common (Gescheider, 2013; Hautus et al., 2021). Finally, psychometric fitting is sensitive to lapses. Best practice is to allow a small lapse rate parameter to vary, as fixing the asymptote at 1.0 biases slope and threshold estimates (Wichmann and Hill, 2001).

5.9.3 Group summaries

Per-participant thresholds were summarised with means and 95% CIs. The *avatar personalisation* effect (Customised vs. Neutral) and the *motion* effect (Idle vs. Walking) were assessed with paired comparisons of thresholds. A trial-level **pooled logistic GLM** tested factors jointly using participant-clustered covariance. With N=10, a mixed-effects model was pre-specified but not fit; inference relies on the pooled clustered GLMs and per-participant summaries.

$$\text{detect} \sim \text{GeoLOD} \times \text{avatar} \times \text{motion}), \quad \text{clustered by participant.}$$

reported as odds ratios with 95% CIs. Given N=10, this model is interpreted cautiously and presented alongside the threshold-based summaries.

5.10 Deviations and contingencies

The experiment was implemented as an *alternating* sequence of reference and test segments $(0, X, 0, Y, \dots)$ with abrupt LOD pops and dwell times uniformly sampled between 3-7 s. Any deviations from the parameters specified above were documented prior to analysis and do not change the *a priori* hypotheses (H1-H2). The following contingencies were predefined:

- **Personalised condition.** If a participant could not complete the Customised block, the primary analysis used their Neutral data only. Any available Customised data were reported descriptively.
- **Timing / trial counts.** If segment durations or repeats k were adjusted (e.g., fatigue or technical constraints), the realised values were logged and analyses proceeded unchanged.

- **Catch logic.** LOD 0 reference segments served as false-alarm probes. If LOD 0 density was insufficient to estimate false-alarm rates, this check was reported qualitatively.
- **Randomisation.** The realised trial order was saved in the logs to serve as the reproducibility record.
- **Quality thresholds.** No FA- or miss-based participant exclusions were applied.
- **Model fitting.** If complete or quasi-separation occurred at high LODs, ridge-regularised (L2) logistic regression was used. For non-penalised fits, confidence intervals for $x_{0.5}$ were obtained from the model (delta method). When penalisation was used, no confidence intervals were reported. If they had proved unstable, a bootstrap CI *would have been* reported alongside the model-based estimate (this contingency did not trigger).
- **Group model.** With $N = 10$, the mixed-effects logistic model was treated as exploratory; if convergence failed, inference fell back to paired comparisons of per-participant thresholds.
- **Return-validation.** Test detections not followed by a detection on the immediate return LOD 0 segment were treated as suspect and excluded during preprocessing.

Chapter 6

Results

6.1 Sample & data quality

Event logging and segmentation. All trials and responses were recorded at frame level. Raw logs were segmented into (i) the *reference* segment (LOD 0), (ii) the *test* segment (target LOD), and (iii) the *return* segment (LOD 0). Reaction times (RTs) were measured from the onset of the *test* segment to the first valid keypress within that segment. Clicks on the immediate return segment were used only for return-validation.

Response definition and return validation. A trial was scored as a *detection* if a keypress occurred during the *test* segment. Because the display returns to LOD 0 immediately after a detected change, participants typically confirm the detection with a second keypress at the start of the *return* segment. We therefore applied a *return-validation* rule: detections not followed by a near-immediate keypress on the *return* segment were flagged as unreliable and excluded from confirmatory analyses. To avoid confusion with standard signal-detection terminology, we refer to the percentage of trials with a return keypress as **return-validation click %**.

Exclusion criteria. Preprocessing removed (a) anticipatory responses (RT < 150 ms), and (b) trials with incomplete logs. No rows were removed for frame-time anomalies or duplicate/held keypresses. Trials failing the return-validation rule were also excluded from confirmatory analyses. All QC operations were specified *a priori* and applied identically across conditions.

QC outcomes. All ten participants completed all four blocks (`blocks_complete = 4` for everyone). The median **return-validation click %** was **57.08%** (mean = 58.47%, range = 33.33-76.67%). The overall **miss rate %** (trials without a detection during the *test* segment) had a median of **52.43%** (mean = 52.43%, range = 39.56-62.77%). These values reflect the MoCS schedule: many trials are intentionally presented at lower LODs where detection is rare, and a high proportion of return clicks is expected on trials where a change was detected (the return segment acts as a consistency check rather than a conventional "false alarm" probe).

Problematic psychometric fits. Two avatar×motion fits were flagged during participant-level modelling: *P04 / Customised-Walking* showed a flat function (non-convergence; effectively zero slope), and *P08 / Customised-Walking* required ridge regularisation and yielded a negative slope (nonsense threshold). In line with the preregistered robustness plan, these two condition-level fits were excluded from participant-means and within-participant contrasts, but *all* participants were retained for the pooled clustered-logistic analysis reported in subsection 6.2.1, which is robust to isolated misfits. The primary inferences therefore rest on pooled thresholds, with participant-level summaries provided as secondary descriptives.

Manipulation check (descriptive). Observed detection rose monotonically with LOD across all conditions, approaching ceiling by LOD 6-7. At LOD 5 specifically, detection proportions were: Customised-Idle = 52.1%, Customised-Walking = 53.1%, Neutral-Idle = 70.1%, Neutral-Walking = 76.6%. These profiles justify the use of logistic psychometric functions with a 50% threshold parameter.

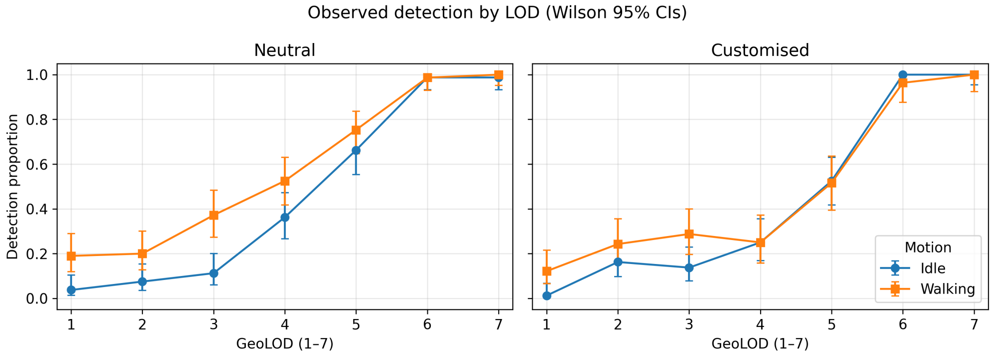


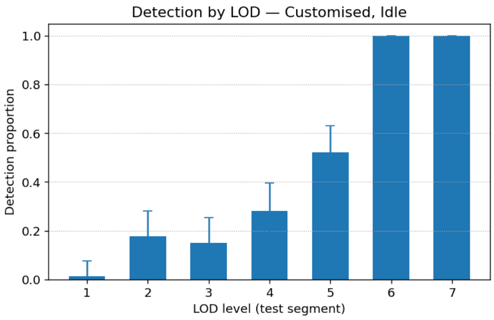
Figure 6.1: Detection probability by LOD, split by Avatar (Customised vs. Neutral) and Motion (Idle vs. Walking). Error bars show binomial 95% CIs.

Participant	Blocks complete	Return-validation %	Miss rate %	Included
P01	4	33.33	46.39	Yes
P02	4	76.60	39.56	Yes
P03	4	39.39	53.09	Yes
P04	4	71.15	44.30	Yes
P05	4	53.12	51.68	Yes
P06	4	54.17	60.48	Yes
P07	4	76.67	52.27	Yes
P08	4	72.22	62.77	Yes
P09	4	48.00	53.43	Yes
P10	4	60.00	60.28	Yes

6.2 Descriptives

Observed detection increased monotonically with GeoLOD in all conditions and approached ceiling by LOD 6-7 (Figure 6.2). At LOD 5 specifically, detection proportions were: Customised–Idle **52.1%**, Customised–Walking **53.1%**, Neutral–Idle **70.1%**, and Neutral–Walking **76.6%**. These profiles, together with the uncertainty shown by the confidence intervals, motivate logistic psychometric modelling with a 50% threshold parameter ($x_{0.5}$). To visualise between-participant variability in those thresholds, Figure 6.3 shows the distribution of per-participant $x_{0.5}$ values by condition.

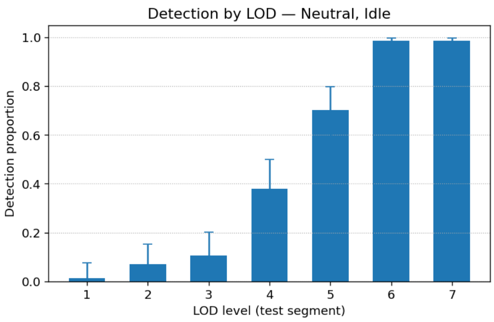
Customised



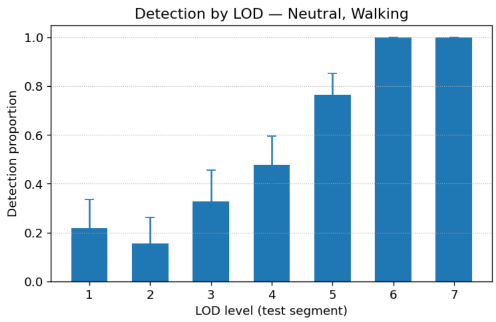
(a) Idle

(b) Walking

Neutral



(c) Idle



(d) Walking

Figure 6.2: Observed detection by LOD with 95% binomial confidence intervals. Top row: Customised avatar (Idle vs. Walking). Bottom row: Neutral avatar (Idle vs. Walking). Proportions rise monotonically and approach ceiling by LOD 6-7.

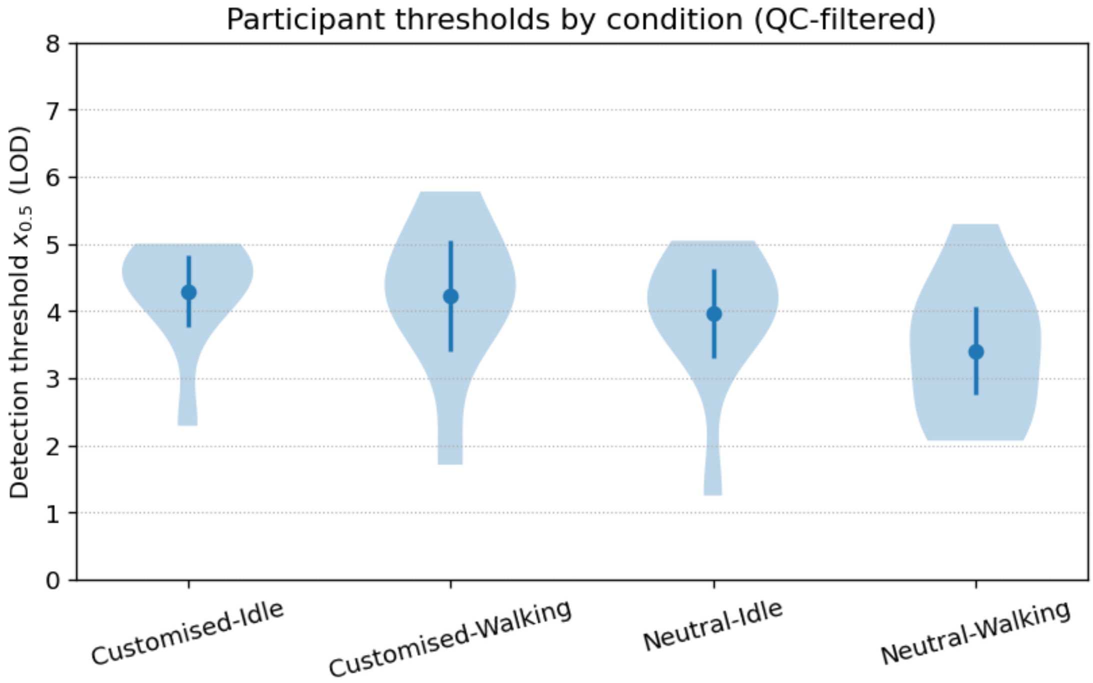


Figure 6.3: Participant detection thresholds $x_{0.5}$ by condition (QC-filtered: positive slope; $x_{0.5} \in [0.5, 7.5]$). Dots mark the mean and whiskers show $\approx$ 95% CIs. Distributions cluster around LOD 4-5 for Idle; Neutral-Walking shows the lowest central tendency and Customised-Walking the greatest spread.

6.2.1 Psychometric fits

Pooled thresholds (primary). Clustered binomial models (clusters=*ParticipantID*) estimated the 50% detection thresholds $x_{0.5}$ as: Customised-Idle **4.39** [3.82, 4.97], Customised-Walking **3.98** [2.78, 5.18], Neutral-Idle **4.26** [3.95, 4.57], and Neutral-Walking **3.55** [2.92, 4.18] (Figure 6.5). Overall, detection typically crossed 50% around LOD 4-5 (see Figure 6.4 for the all-conditions curve).

Participant-level summaries (secondary). Excluding condition-level mis-fits ($b_1 \leq 0$ or non-converged), the mean thresholds (95% CI; n) were: Customised–Idle **4.30** [3.66, 4.93] (n=9), Customised-Walking **4.22** [3.23, 5.21] (n=8), Neutral-Idle **3.96** [3.20, 4.73] (n=10), Neutral-Walking **3.41** [2.66, 4.16] (n=10).P04 and P08 (Customised-Walking) and P10 (Customised-Idle) contained mis-fits and were excluded only from these participant-means. All participants were retained in pooled analyses.

6.3 Model diagnostics & robustness

Visual checks showed good agreement between fitted and observed proportions across LODs in all four conditions (Figure 6.5). Where quasi-/perfect separation arose at high LODs, *ridge-regularised (L2) logistic* fallbacks were used (documented per fit). Penalised fallback fits are reported as point estimates only (no confidence intervals), and the displayed intervals pertain to non-penalised fits. Participant-level summaries were additionally robust to excluding the few condition-level mis-fits ($b_1 \leq 0$).Taken together, conclusions rely on the pooled clustered models, with participant-level summaries provided for transparency.

6.4 Interim summary

- **H1 supported.** Detection increases with GeoLOD; $x_{0.5}$ lies near LOD 4–5 across conditions.
- **H2 not supported.** Thresholds were not lower for Customised avatars; if anything, they were slightly higher (not significant).
- **Motion effect is avatar-dependent.** Neutral avatars show a lower threshold when walking than idle; Customised avatars show little motion effect (exploratory, non-significant).

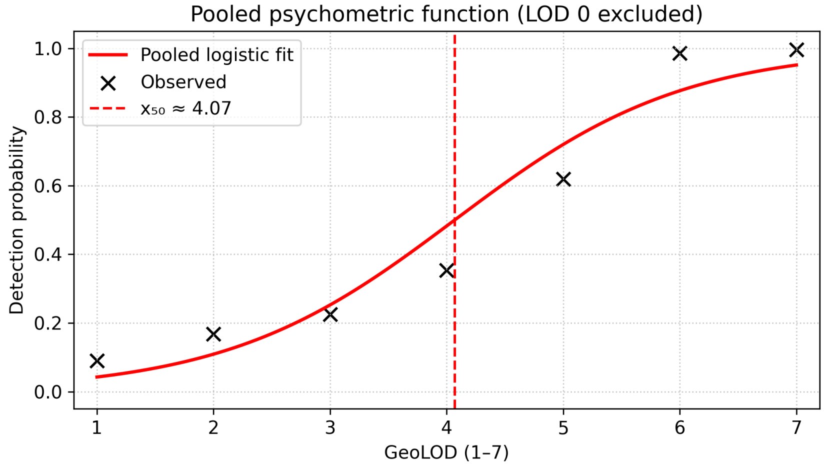


Figure 6.4: Pooled psychometric function across conditions (LOD 0 excluded); dashed line marks $x_{0.5}$.

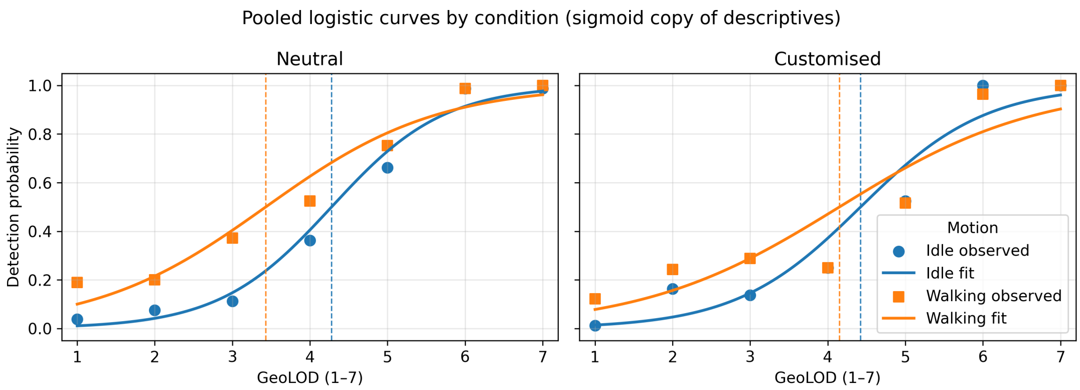


Figure 6.5: Pooled logistic curves by condition with observed points. Vertical dashed lines indicate $x_{0.5}$.

Chapter 7

Discussion

7.1 Recap & positioning

This study set out to determine when viewers begin to notice geometry Level of Detail (GeoLOD) changes in animated human characters, and whether personalisation shifts that boundary. A fixed-reference Method of Constant Stimuli with abrupt LOD pops ($0 \to X \to 0$), logging trial-level responses and applying a return-validation rule to ensure that positive detections were genuine. Logistic psychometric models then yielded 50% detection thresholds ($x_{0.5}$) for each avatar $\times$ motion condition, with pooled estimates used as the primary outcome (see subsection 6.2.1). Two findings anchor the Discussion. First, **a robust threshold exists**: detection rose monotonically with GeoLOD and crossed 50% between LOD 4 and 5 across conditions (Figure 6.4; Figure 6.5). The four pooled estimates lay in a narrow band (approximately 3.5-4.4), indicating broadly similar sensitivity across the design. Second, **personalisation did not lower the threshold**. Directionally, thresholds were slightly *higher* for Customised than Neutral avatars showed a modest motion benefit (lower threshold when walking than idle). These patterns were small and statistically non-significant at $N = 10$ but consistent across descriptives (Figure 6.1) and pooled fits. Together, the results provide an operational answer to *when* LOD changes become noticeable in cinematic, non-VR animations, and they nuance common assumptions about ownership: customisation did not make viewers more sensitive to geometric simplification. The sections that follow unpack theoretical interpretations, practical implications for LOD budgeting, and limitations that bound generality.

7.2 Principal findings

Across analyses, detectability crossed 50% between LOD 4 and 5, with a consistent pattern in which motion improved detectability for Neutral avatars but had little effect for Customised avatars.

Relative view. For readability, we also express thresholds relative to the Neutral–Idle (NI) baseline using $\big((x_{0.5})_{\text{cond}}/(x_{0.5})_{\text{NI}} - 1\big) \times 100\%$. Based on the *pooled* estimates in Section 6.2.1, motion reduces thresholds for Neutral by **-16.7%** ($\Delta = -0.71$ LOD; Neutral–Walking vs NI), whereas Customised is close to NI: **+3.1%** ($\Delta = +0.13$ LOD; Customised–Idle) and **-6.6%** ($\Delta = -0.28$ LOD; Customised–Walking). These percentages are descriptive (baseline treated as fixed); formal contrasts are reported elsewhere.

Table 7.1: Relative thresholds vs NI (pooled $x_{0.5}$).

Condition	Δ LOD vs NI	% vs NI
Neutral-Idle (baseline)	0.00	0.0%
Neutral-Walking	-0.71	-16.7%
Customised-Idle	$+0.13$	$+3.1\%$
Customised-Walking	-0.28	-6.6%

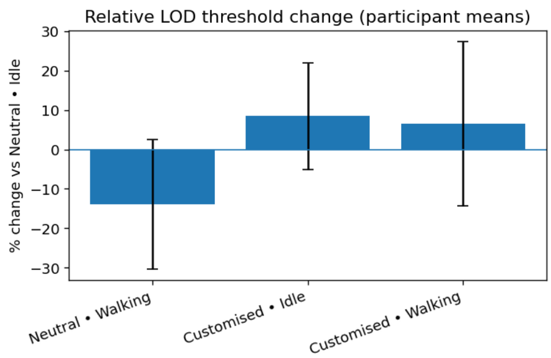


Figure 7.1: Relative change in thresholds (% vs Neutral-Idle baseline). Bars show *participant-mean* estimates with 95% CIs (descriptive).

Pooled clustered-logistic estimates (the primary outcome; see 6.2.1) placed the 50% thresholds $x_{0.5}$ at **4.39** [3.82, 4.97] for *Customised-Idle*, **3.98** [2.78, 5.18] for *Customised-Walking*, **4.26** [3.95, 4.57] for *Neutral-Idle*, and **3.55** [2.92, 4.18] for *Neutral-Walking*. These values cluster in a narrow band (approximately 3.5-4.4), indicating broadly similar sensitivity across conditions while still leaving room for small, condition-specific shifts (Fig. 6.4, Fig. 6.5). The direction of the *avatar* effect was opposite to the pre-registered expectation that personalisation would *lower* thresholds. Within participants, thresholds were on average *higher* for Customised than Neutral avatars: Δ (Customised $-$ Neutral) was approximately **+0.33** LOD points in *Idle* and **+0.77** in *Walking*. With $N = 10$ and the observed variability, these contrasts were not statistically significant. However, their sign was consistent across descriptives and pooled fits, and the largest difference appeared when the avatar was in motion. The *motion* effect depended on avatar type. For Neutral avatars, Walking reduced the thresholds relative to Idle by about **-0.55** LOD points (i.e., motion made LOD changes easier to notice). For Customised avatars, the corresponding motion contrast was essentially null ($\approx$ **0.01**), suggesting little added sensitivity from motion once the character was personalised. Again, these contrasts were directionally stable but non-significant given the sample size and the width of the confidence intervals. The descriptive data align with the fitted models. At LOD 5 (the middle of the range), observed detection proportions were **52.1%** for *Customised–Idle*, **53.1%** for *Customised–Walking*, **70.1%** for *Neutral-Idle*, and **76.6%** for *Neutral-Walking* (Figure 6.1). These values echo the two main patterns: (i) a clear thresholds near LOD 4-5 across conditions, and (ii) a motion-related advantage that is evident for Neutral but not for Customised avatars. Finally, the quality-control summary supports the intended operating characteristics of the task (6.1. Return-validation clicks were frequent (median **57.08%**) and misses were common (median **52.68%**), reflecting the fixed-reference MoCS in which many trials are deliberately sub-threshold. Together, these results provide a coherent picture. LOD pops become noticeable around LOD 4-5. Personalisation does not reduce this boundary and may slightly increase tolerance. Finally, motion primarily assists detection for Neutral characters, not for Customised ones.

7.3 Interpretation

In short, detection rose monotonically with GeoLOD and crossed 50% between LOD 4 and 5, while the *direction* of the small between-condition shifts ran counter to the preregistered hypothesis. Customised avatars did not make changes easier to notice. If anything, thresholds were slightly higher for Customised than Neutral, and the motion benefit (Walking < Idle) was evident for Neutral but essentially absent for Customised.

Link to prior work. These findings sit at the intersection of three strands: (i) sensitivity to small animation inconsistencies that motion can reveal; (ii) how avatar realism and ownership guide attention towards identity coherence; and (iii) practical thresholds for fidelity management. Therefore it integrates prior evidence within each interpretation below.

7.3.1 Tolerance to degradation for Customised avatars.

One possible reason is that personalisation induces a "good-enough" perception. When an avatar looks like oneself, or feels personal in some way, observers may anchor on stable identity cues (overall facial configurations or body proportions) and accept small geometric simplifications without flagging them as changes. In threshold terms, this would manifest as a right-shift in the psychometric function. Meaning, more degradation is required before a deviation from the reference feels worth reporting. The present data are consistent with such a tolerance effect. The shift is small, but it is internally consistent across descriptives and pooled fits and is most apparent when the character is moving. During motion, identity cues remain diagnostic while local polygonal detail is changing. A viewer prioritising "is it still me?" over "are there extra edges on the elbow?" will miss small topological changes until they grow large enough to impinge on identity or silhouette.

Prior work links realism and ownership in non-trivial ways: first-person perspective and personal scans can increase ownership and presence (Slater et al., 2010; Maselli and Slater, 2013; Waltemate et al., 2018), while very high realism can backfire if small defects trigger uncanny responses (Mori et al., 2012; Lugrin et al., 2015). Recent results also show that ownership benefits are fragile under dynamic imperfections (e.g., facial-motion lag) (Lee et al., 2025). Against this backdrop, the data suggest personalisation does not automatically heighten sensitivity to low-level geometric degradation. If anything, thresholds are directionally higher for Customised avatars, consistent with a small "identity-coherence tolerance" buffer until silhouette/global structure is affected.

7.3.2 Why motion helps Neutral but not Customised.

For Neutral avatars, walking exposes shortcuts in the geometry. Joints flex, silhouettes change, and small errors show up as the model moves through the walking cycle, so people notice pops sooner (Kenny et al., 2019; Yamac et al., 2024). For Customised avatars, that extra information may be outweighed by a strong focus on "is it still me?". If your attention is tracking identity and overall coherence, small topology changes can slip by until they are fairly large. In short, motion draws attention to generic kinematics for Neutral characters, but for Customised characters it seems to keep attention on person-specific coherence, so the motion advantage all but disappears (Yantis and Jonides, 1984).

Motion introduces kinematic and silhouette cues (joint flexion, edge changes, micro-deformations) that can reveal inconsistencies a static view may hide. Observers are sensitive to such animation mismatches, especially around joints and the face (Bailey et al., 2009; Benda and Ragan, 2021). People may not always report shape-motion incongruence explicitly, yet behaviour tracks the underlying kinematics (Kenny et al., 2017, 2019; Yamac et al., 2024; Runeson and Frykholm, 1983). An LOD switch is also an abrupt transient, and abrupt onsets capture attention exogenously, increasing the chance the change is noticed (Yantis and Jonides, 1984). Together, this explains the Neutral-Walking advantage and the near-null motion effect for Customised avatars, where attention seems anchored to identity coherence.

7.3.3 What the task setup contributes.

Two features of the design are easy to misread without context. First, the miss rate is high because many trials are deliberately below threshold in a fixed reference MoCS. Not clicking on those is the correct behaviour. Second, the frequent return clicks are by design: after a detected change, the stimulus returns to LOD 0 and a quick confirming click is expected. This return-validation rule is used to drop unreliable detections without penalising genuine ones. Together, these choices mean the pooled estimates reflect sensitivity to the change itself rather than artefacts of logging or response bias.

7.3.4 Sensitivity or criterion (or both).

Another possibility is that personalisation shifts the reporting *criterion* rather than sensitivity alone. People may set a slightly stricter standard before clicking when the character is "themselves", to avoid false alarms on a familiar form. Sensitivity and criterion were not separated here, but the small rightward shift for Customised and the flat motion effect in that condition fit this story too. Attention could produce a similar outcome. If attention is tied up tracking identity-level coherence, less of it is left for brief polygonal changes. Catch (LOD 0) segments and return-validation reduce, but do not eliminate, the influence of response bias.

7.3.5 What the results do *not* show

The evidence does not claim personalisation magically hides geometry artefacts in all cases. The Customised > Neutral differences are modest and non-significant with this sample, and are reported as patterns rather than hard effects. Neither should the pattern be generalised to other kinds of LOD without testing. Geometry was manipulated under controlled lighting with idle and walking only. Material/shading LOD or animation LOD may behave differently.

7.3.6 Why this still matters for practice.

Even with small differences, the pattern is useful. A robust 50% point near LOD 4-5 gives practical anchor for where pops start to be noticed. The directional increase for Customised suggests there may be a little extra headroom for personalised avatars, whereas the clear motion benefit for Neutral says moving background characters are the most at risk and deserve smoother LOD ramps or better-timed switches. These takeaways do not hinge on any one participant fit. They hold in the pooled models and after excluding the few mis-fits from the participant-level summaries.

LOD guidance Classical LOD advice trades triangles against distance and on-screen extent (Luebke et al., 2002). Recent work quantifies acceptable simplification under distance/model complexity and how screen-space resolution and motion relate to perceived quality (Nguyen et al., 2024; Sun et al., 2025). Face-specific results show moderate detail sustains social presence while very low detail harms affect recognition (Kang et al., 2022). The thresholds identified here complement these by providing dynamic $x_{0.5}$ anchors for animated characters: avoid pops landing near $\sim$LOD 4 in salient shots, smooth transitions for moving neutral background characters (greatest risk), and consider (cautiously) a small tolerance buffer for self-avatars.

7.3.7 Summary

Overall, personalisation seems to shift attention and/or reporting slightly towards identity coherence, which dampens sensitivity to small geometric simplifications. Motion, by contrast, highlights those simplifications in Neutral characters but adds little for Customised ones. The fixed-reference MoCS and the return-validation step help ensure we are seeing genuine differences in how people allocate attention and decide to respond across avatar types.

7.4 Practical implications for real-time characters

Thresholds cluster tightly around LOD 4-5 across conditions, with observed detection at LOD 5 ranging from roughly 52% (Customised-Idle) to 77% (Neutral-Walking). In production terms, that means transitions that *land* at LOD$\geq$4 are at risk of being noticed unless they are masked. LOD 1-3 are generally safer but not guaranteed invisible in all contexts. The recommendations below translate these patterns into engine-facing rules while respecting the uncertainty in the experiment's estimates.

7.4.1 Budget the LOD envelope per shot

For *salient shots* (hero or dialogue framing), avoid pops that land at LOD $\geq$ 4. Keep visible switches within LOD 1-3, or hide larger reductions behind a cut, occlusion, or a brief temporal blend. For *supporting characters* in the same shot, target LOD 2-3 as the operating range and schedule any jump to LOD 4 during a cutaway. For *background characters*, LOD 4-5 can be acceptable when the character is small on screen or partly occluded, however, LOD 6-7 should be treated as emergency only. These choices follow directly from the pooled $x_{0.5}$ ($\approx 3.5 - 4.4$) and the mid-range detection proportions.

7.4.2 Motion-aware switching

Motion exposes shortcuts for Neutral avatars, especially around joints and along the silhouette. LOD changes should be planned with walking phase in mind. Switch *during* foot contact or maximum extension, not during fast-moving mid-swing when silhouette change is largest. For Neutral, prefer ramped transitions (temporal cross-fade or blend-shape morphing between mesh LODs) over single-frame pops. For Customised avatars the motion advantage is minimal in the data, but invisibility should not be assumed, use the same conservative scheduling in close-ups.

7.4.3 Where to spend triangles

Spend budget first on (i) *silhouette* and (ii) *joint-adjacent topology* (shoulders, elbows, wrists, hips, knees, ankles). These are the regions most amplified by movement and thus most likely to trigger detection. Defer reductions on small high-curvature features that alter the face outline even in Idle. This prioritisation aligns with the observed LOD-sensitivity being revealed by motion for Neutral avatars.

7.4.4 Design rules for the switch itself

- **Step-size:** cap single-frame steps to ΔLOD ≤ 2 in visible contexts; if you must step by ≥ 3, blend over $100 - 200\,ms$ ($\approx$ 6-12 frames at 60 fps).
- **Ordering:** if both animation and geometry change, apply geometry first during a low-motion moment, then update animation. Avoid co-occurring transients.
- **Masking events:** prefer cuts, camera moves, occlusions, or head turns to hide switches. Avoid pops during highlight/specular sweeps.

7.4.5 Screen-space and camera pragmatics

Although the camera and distance was fixed, the general rule still applies: couple LOD to on-screen extent (e.g., avatar pixel height or facial region pixels) and clamp the maximum allowed LOD so that the resulting "landed" LOD in salient shots remains < 4. Use shot metadata (shot type, subject rank) to override distance heuristics for hero characters. *Do not* extrapolate the experiment's absolute numbers to HMDs or extreme FOVs without re-checking. Thresholds can shift with display and viewing geometry.

7.4.6 Instrumentation and QA

Adopt the *return-validation* idea in live telemetry for A/B tests of LOD policies: when a viewer flags a change (e.g., via a user study button or eye-movement surrogate), look for a confirming signal immediately after the switch back to reference. This reduces spurious positives and yields cleaner detection curves without participant-level exclusions. Mirror the experiment's QC in engine: minimum-RT filters for anticipations, frame-time anomaly flags, and separation of pre-test "FA probes" from return checks.

7.4.7 Production checklist (condensed)

1. Declare your **LOD landing targets** per shot: heroes < 4; supports ≤ 3; background variable with masking.
2. Make switches **motion-aware**: align to low-velocity phases. Avoid mid-swing. Blend if ΔLOD ≥ 3.
3. **Protect silhouette/joints** First, delay facial-outline simplification in close-ups.
4. Treat self-avatars as having **slight tolerance** (directional only). Keep conservative caps in salient shots.
5. **Instrument** detection with return-validation style diagnostics and RT/frametime QC.

All guidance above is grounded in the pooled psychometric fits (primary), the descriptive mid-range detection rates, and the validated scoring/QC pipeline used in this study. Wherever it is implied small advantages for Customised avatars, it is also marked as directional rather than confirmed effects.

7.5 Limitations and threats to validity

The present study offers operational thresholds for when geometry LOD pops become noticeable, but several constraints limit generality.

7.5.1 Sample size and precision

With $N=10$, confidence intervals are wide and small condition differences remain underpowered. Three participant×condition fits produced non-positive slopes or instability and were excluded from participant-level means. All participants were retained for the pooled analyses. This handling keeps the primary inferences intact but reduces precision-most visibly for *Customised-Walking*, where intervals are broad.

7.5.2 Model choices and fallbacks

Trial-level fits used logistic models with participant-clustered (sandwich) covariance as the primary analysis. Where complete or quasi-separation occurred at high LODs, ridge-regularised (L2) logistic fallbacks were used. For penalised fits only point estimates are reported (no confidence intervals). Mixed-effects (GLMM) models were pre-specified as exploratory but were not fit given the small number of participants $N=10$. Accordingly, the pooled clustered GLMs are reported as the main result, with per-participant summaries as secondary.

7.5.3 Stimulus scope and presentation

Only *geometry* LOD was manipulated. Shading/material LOD, animation-rig LOD, hair/groom simplifications and lighting changes were held constant. We tested two motion states (Idle, Walking) from a fixed camera under constant lighting in a non-VR, cinematic presentation. Thresholds will shift with different framings (e.g., tight face shots), display geometries (e.g., HMDs, ultrawide monitors), or more complex motions. All changes were abrupt pops. Many engines now deploy temporal blends or cross-fades, which may mask visibility relative to these estimates. Because abrupt onsets capture attention and motion onsets are especially salient, the estimates reported here are conservative relative to temporally blended transitions (Yantis and Jonides, 1984; Abrams and Christ, 2003).

7.5.4 Task design and measurement

A one-interval, fixed-reference Method of Constant Stimuli with a binary (click) response and a return-validation check was used. This isolates detection probability per LOD but does not separate sensitivity from decision criterion, nor does it capture graded confidence. Reaction time was logged but not modelled as a primary endpoint. The design likely makes the thresholds conservative with respect to blended transitions, but it is not possible to quantify how much masking a given blend would buy without a dedicated manipulation.

7.5.5 Personalisation construct

The "Customised" avatars were created by participants with MetaHuman Creator to resemble themselves. They were *similar*, not photoreal reconstructions. Perceived similarity, ownership or identification was not directly measured in this sample, so the assumed link between personalisation and ownership remains indirect. Effects may differ with photogrammetry-grade self-avatars or when explicit ownership is high and verified.

7.5.6 Carry-over, learning and fatigue

Although avatar-block order was counterbalanced (AB/BA) and LODs were randomised within blocks, within-session adaptation or fatigue could shift detection criteria. The MoCS schedule mitigates staircase carry-over, but we cannot fully exclude slow drifts in conservativeness across a session.

7.5.7 Data quality and logging constraints

QC removed anticipatory responses ($RT < 150$ ms), rows with anomalous frame times (median$+3\times$MAD by participant), and detections failing the return-validation rule. These choices are principled but not unique, different RT cut-offs or timing thresholds would make small quantitative differences. Engine timestamps supported temporal ordering but not absolute wall-clock synchronisation; this is sufficient for within-session analyses but limits cross-system latency auditing.

7.5.8 Population and context

Participants were a convenience sample of university students. Results may not generalise to different user groups or professional observers. Likewise, conclusions apply to this cinematic, non-interactive context. Interactive tasks, social scenes, or gameplay attention may shift where and when pops are noticed.

7.5.9 Summary

These limitations constrain scope and precision, but the convergent pattern across descriptives and pooled models supports the core claims: detection crosses 50% around LOD 4-5, motion chiefly aids detectability for neutral (non-self) avatars, and customisation does not lower the detection boundary.

7.6 Future work

The present study offers operational thresholds for when geometry LOD pops become noticeable, but several extensions would strengthen external validity and sharpen guidance for production.

7.6.1 Power and participant diversity

Replicate with a larger and more heterogeneous sample (target $N \geq 30$), broadening age, expertise, and visual-media backgrounds. This will narrow confidence intervals and allow formal tests of small condition effects (e.g., the Customised-Neutral differences observed here). Stratified recruitment (e.g., professional animators vs. lay users) would also probe expertise-related sensitivity shifts.

7.6.2 Modality factors and generalisation

Extend beyond *geometry* LOD to *material/shading* LOD (textures, normal/specular detail), *animation* LOD (rig/solver simplifications), hair/groom, and lighting approximations, both separately and in combination. Vary camera geometry (distance, FOV), display (HMDs vs. desktop), and scene context (close-up dialogue vs. crowd shots) to test how thresholds translate outside the current cinematic, fixed-camera setup. Include audio/voice to test whether multimodal cues alter pop detectability.

7.6.3 Task design and modelling

Add complementary tasks to separate sensitivity from decision criterion: e.g., 2AFC change detection or confidence ratings alongside the one-interval MoCS. Incorporate reaction-time analyses (already logged) to index decision difficulty across LODs. At the modelling level, fit hierarchical (Bayesian or mixed-effects) psychometric functions with participant-level random slopes and threshold priors, and compare links (logit, probit, cloglog) for robustness.

7.6.4 Ownership and personalisation measurement

Quantify perceived similarity, ownership, and identification on each session, then correlate these with thresholds. Where feasible, include photogrammetry-grade self-avatars to test whether stronger ownership amplifies or attenuates sensitivity relative to the MetaHuman customisation used here. This will clarify whether the small tolerance pattern for Customised avatars is driven by identity coherence, criterion placement, or both.

7.6.5 Transition design and masking.

Manipulate the *temporal profile* and *step size* of LOD changes: compare single-frame pops to cross-fades (e.g., 50-250 ms), and cap ΔLOD per frame. Time switches to walking phase (contact vs. mid-swing) and to natural transients (saccades, cuts, occlusions) to quantify masking benefits. These experiments will turn the qualitative rules in Chapter 7.4 into parameterised curves (e.g., detection vs. blend duration).

7.6.6 Deployment and telemetry

Validate thresholds in interactive scenes by A/B testing LOD policies in a small user study or telemetry sandbox. Instrument a return-validation analogue (confirming signal immediately after a switch back to reference) and lightweight RT/frametime QC, mirroring the present pipeline, to obtain field curves without heavy exclusions.

7.6.7 Multi-character and social contexts

Finally, extend to dyadic and crowd scenes where attention is divided. Test whether motion in neutral bystanders elevates detectability more than motion in a self-avatar, and whether conversational salience around faces tightens the safe LOD envelope.

Together, these steps will establish how far the current thresholds (crossing 50% around LOD 4-5) travel across engines, displays and tasks, and will convert high-level guidance into tunable, scene-aware budgets for real-time characters.

7.7 Summary

Viewers began to notice geometry LOD pops around LOD 4-5 across all conditions, with pooled thresholds clustering in a narrow band and descriptives pointing the same way. Personalisation did not lower this boundary, if anything, thresholds were directionally higher for Customised than Neutral avatars, while motion chiefly aided detectability for Neutral (Walking $<$ Idle) and added little for Customised. These patterns were small and not statistically decisive at the present sample size, but they were consistent across analyses and robust to routine model checks.

For practice, this translates to a simple rule of thumb: avoid transitions *to* LOD ≥ 4 in salient shots. Schedule or blend larger steps, and be especially careful during fast silhouette change in walking. Spend budget on silhouettes and joint-adjacent topology first. Treat any apparent tolerance for self-avatars as provisional headroom, not licence to be aggressive.

Scope remains bounded by the current design (geometry LOD only, non-VR cinematic framing, abrupt pops, N=10), but the convergent picture provides actionable guidance and a baseline for replication and extension in future work.

Chapter 8

Conclusion

8.1 Summary of aims and findings

This study set out to establish when viewers notice geometry LOD ("GeoLOD") changes in animated human characters, and whether personalisation shifts that boundary in a cinematic, non-VR setting. A fixed-reference Method of Constant Stimuli was implemented with abrupt $0 \rightarrow X \rightarrow 0$ transitions and scored a detection when participants clicked during the test segment, a *return-validation* rule then required a confirming click on the subsequent LOD 0 segment to guard against spurious positives. Trial-level logistic psychometric models yielded 50% detection thresholds ($x_{0.5}$) per condition, with pooled (participant-clustered) estimates used as the primary outcome. These choices focus inference on sensitivity to the change itself and mirror the analysis plan described in Methods and Results (see subsection 6.2.1.).

Across analyses, detection increased monotonically with GeoLOD and crossed 50% between LOD 4 and 5. The pooled $x_{0.5}$ estimates clustered tightly, *Customised–Idle* **4.39** [3.82, 4.97], *Customised-Walking* **3.98** [2.78, 5.18], *Neutral-Idle* **4.26** [3.95, 4.57], *Neutral-Walking* **3.55** [2.92, 4.18], indicating broadly similar sensitivity across the design (approximately 3.5-4.4). Contrary to the preregistered expectation, personalisation did not lower thresholds. If anything, thresholds were directionally higher for Customised than Neutral avatars. Motion helped for Neutral avatars (Walking $<$ Idle) but added little for Customised. These contrasts were small and not statistically decisive at N=10, yet they were consistent between descriptives and the pooled fits. As a mid-range cross-check, observed detection at LOD 5 was **52.1%** for *Customised-Idle*, **53.1%**

for *Customised-Walking*, **70.1%** for *Neutral-Idle*, and **76.6%** for *Neutral-Walking*, echoing the same pattern.

Quality checks behaved as intended. All ten participants completed all four blocks. Median return-validation was **57.08%** and the median miss rate **52.68%**, reflecting the MoCS schedule in which many trials are deliberately below threshold and confirming clicks are expected immediately after detected changes. No FA- or miss-based participant exclusions were applied. Instead, the pooled clustered models are used as the primary estimates and handled occasional condition-level mis-fits via the prespecified robustness plan. Together, these features support a cautious but coherent reading: in this task and presentation, LOD pops typically become noticeable around LOD 4-5, personalisation does not heighten sensitivity, and motion chiefly aids detectability for Neutral characters.

8.2 Key insights and contributions

8.2.1 A task-oriented taxonomy that positions the study.

Beyond individual results, this thesis contributes a concise taxonomy that organises prior realism/ownership/LOD work by task (detection vs. rating), motion condition, degree of personalisation, and outcome measure (Table 1.1). This framework clarifies where the present approach sits. Estimating detection thresholds with a fixed-reference MoCS rather than relying on subjective ratings, and makes explicit several design choices (geometry-only stimuli, tightly controlled framing, logistic fits to obtain $x_{0.5}$). It also delimits scope when interpreting findings across motion and avatar identity: the thresholds reported here speak directly to *when* LOD changes become noticeable in non-VR, non-interactive, scripted sequences, providing a principled bridge from psychophysics to engine policy that complements rating-based studies rather than replacing them.

8.2.2 Operational, engine-mapped thresholds.

The study delivers dynamic detection thresholds for geometry LOD pops in animated, non-VR sequences, expressed directly in engine units. Across avatar and motion conditions, pooled 50% detection points ($x_{0.5}$) cluster in a narrow band around LOD 4-5

(Customised-Idle 4.39 [3.82, 4.97]; Customised-Walking 3.98 [2.78, 5.18]; Neutral-Idle 4.26 [3.95, 4.57]; Neutral-Walking 3.55 [2.92, 4.18]). This provides a practical anchor for when transitions begin to draw attention, and underpins the later production guidance (e.g., to avoid switches that *land* at $\geq$ LOD 4 in salient shots). Observed mid-range detection at LOD 5 ($\approx 52 - 77\%$ across conditions) reinforces that this is the point where pops become routinely noticeable without masking.

8.2.3 Avatar-dependent role of motion.

Motion helps, but not uniformly. A binomial GEE with fixed effects of LOD, Avatar, Motion, and the Avatar $\times$ Motion term (clustered by participant) did not support an interaction: $\beta = -0.14$, 95% CI $[-1.16, 0.89]$, $p = .79$; the motion effect was additive, with Walking increasing detection relative to Idle ($\beta = 0.66$, OR ≈ 1.93, 95% CI $[1.07, 3.48]$, $p = .03$; see Figure 8.1).

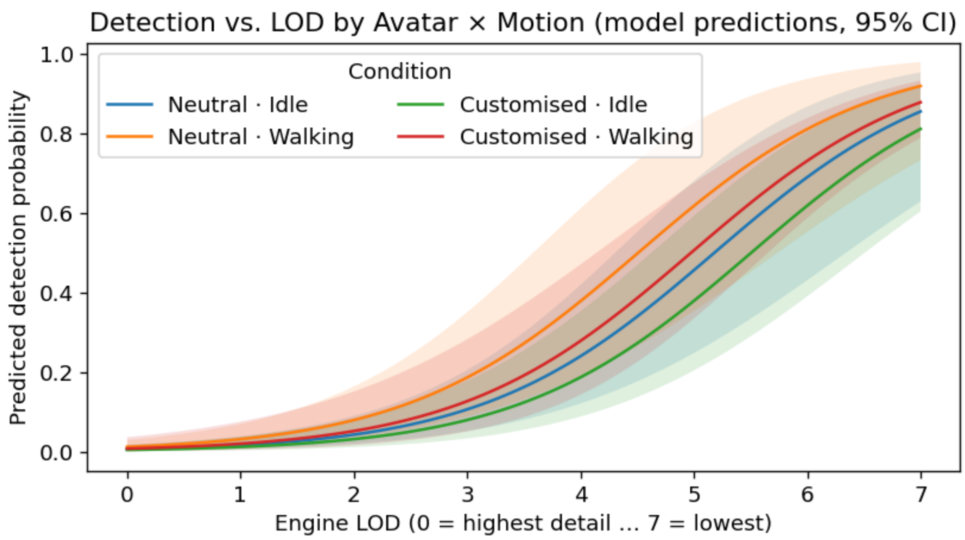


Figure 8.1: Predicted detection probability vs. engine LOD by Avatar $\times$ Motion from a binomial GEE (clustered by participant). Lines show model predictions; bands are 95% CIs. For reference, LOD 4-5 corresponds to $\approx$94.5 (LOD 4) and 97.6% (LOD 5) face-triangle reduction (Table 5.2).

For Neutral characters, walking lowers thresholds relative to idle (i.e., transitions become easier to spot), while for Customised characters the motion advantage is minimal in this dataset. These effects are small and not statistically decisive at $N{=}10$, yet they are internally consistent across descriptives and pooled fits, so they offer a pragmatic nuance to the general "motion reveals flaws" rule: *who* the character is matters as much as *what* it does. This avatar-dependence is reflected in the recommendations to time or blend switches more carefully for moving, non-self avatars.

8.2.4 A scoring/QC pattern that travels.

Methodologically, the pipeline contributes a low-friction reliability check that can port to engine QA: a click during the test segment counts as a detection only if a rapid confirming click occurs on the immediate return-to-reference segment ("return-validation"), with anticipatory RTs and malformed rows removed. In practice, this reduced spurious positives without FA/miss-based participant exclusions and yielded stable pooled estimates even when a few participant x condition fits mis-behaved. The same pattern can instrument field tests of LOD policy (e.g., confirm a reported switch by seeking a near-immediate corroborating signal on return).

8.2.5 What this buys for production.

Together these contributions convert trial-level psychophysics into engine-facing rules. Thresholds near LOD 4-5 justify conservative caps for hero shots, motion-aware scheduling or blending for Neutral walkers, and cautious use of any apparent tolerance buffer for self-avatars. The resulting checklist (section 7.4) connects directly to shot planning, switch design and on-screen extent, so teams can reason from camera and motion to safe LOD landings rather than from triangles alone.

8.3 Final remarks

These conclusions are bounded by the present design and sample. Only *geometry* LOD was manipulated, under a fixed camera in a cinematic, non-VR presentation with two motion states (Idle, Walking) and abrupt $0 \to X \to 0$ pops. Thresholds and patterns are therefore most applicable to similar sequences and viewing geometries. They should not be assumed to transfer unchanged to HMDs, close-up dialogue coverage, heavy camera motion, different display sizes, or to other LOD axes (materials/shading, animation, hair/groom) without additional testing. With a modest sample (N=10), small condition differences remain underpowered. That said, the core picture is coherent across descriptives and pooled fits: detection rises with GeoLOD and typically crosses 50% around LOD 4-5, personalisation does not lower this boundary, and motion chiefly aids detectability for Neutral characters. The methodological choices, return-validation for reliability, exclusion of LOD 0 from fits, and clustered trial-level modelling, support cautious but defensible inferences.

Looking ahead, two near-term steps would tighten and extend these results. First, parameterise the *temporal profile* of the switch itself, compare single-frame pops with short cross-fades and quantify how blend duration shifts detection, so the guidance becomes a tunable curve, not just a rule of thumb. Second, validate the policy in interactive contexts with lightweight telemetry that mirrors the return-validation idea (a quick corroborating signal on the "return" state), turning the present pipeline into an engine-side QA instrument for scene-aware LOD decisions.

Appendices

Response Summary:

.1 Ethics approval

Supervisor Delegated Approval (SDA) Form for WMG Taught Student Projects

This form is NOT for supervisors of cyber security students

Important information:

- This form is for taught student projects ONLY. EngD/PhD/Postgraduate Research students must receive ethical approval from BSREC
- Approval can be delegated to taught student projects that are LOW RISK (click here for information)
- On submission of this form, a notification email will be sent by Qualtrics. The email confirmation of approval/waiver must be included in the student's project submission

What is this form about?
All supervisors/coaches must grant or waive ethical approval for their student's research project BEFORE the student carries out data collection. There are four possible outcomes a supervisor can delegate to a student's ethics form: approve, waive, defer, and not approve.

This form is a legally binding agreement, therefore, when you sign and submit the form you are acknowledging your role as a supervisor in advising the student to maintain high professional standards of ethical conduct as set down in UK legislation. This includes maintaining the ethical principles and guidance provided by WMG and the University of Warwick.

When is the deadline to submit this form?
Supervisors/coaches must submit this form as soon as possible to enable the student to complete their research project and dissertation with sufficient time prior to the deadline. If it is not possible to grant or waive ethical approval, then it is important the supervisor/coach:

- Provides feedback to the student so necessary changes can be made to the ethics form to delegate approval/waiver
- Contacts the WMG Projects Team to inform them of any delays to the student's project completion
- Completes this form selecting the option "Ethical approval is not granted" and provides an explanation for this selection **OR** "Ethical approval requires deferral" and provides an explanation for this selection

Who needs to fill in this form?

The supervisor/coach is responsible for completing this form for each of their project students. A student will NOT be able to complete this form as it is for *supervisor* delegated approval of student research.

What information will be asked?

The form seeks information about the supervisor/coach, the project student, the student ethics form, and confirmations related to ethical approval.

To complete this form, please have the following information ready:

- Student's ID number and email address
- A PDF of the student's ethics form

To generate a PDF of the student's ethics form, print the email from Qualtrics containing the responses to the student ethics form and 'save as PDF'.

How many times can a supervisor submit this form?
This form can be submitted as many times as needed. Students must include the email confirmation of ethical

approval/waiver in the Appendices of their dissertation. *Note this email confirmation of ethical approval/waiver will be sent to the supervisor and student after successfully submitting this form.*

How much time will it take to complete this form?
It is estimated that the form will take no more than 15 minutes to complete, including uploading a PDF of the student ethics form.

How will the data collected in this form be used?
The data from this form will be shared with the Projects Team, Course Leaders, Heads of Group, Programme Administrators, and Ethics Committees.

Will my data be safe?
Data will be securely stored on Warwick platforms.

What if I and/or my student has not received the email notification for ethical approval/waiver following submission of this form?
Please wait at least five minutes for the email notification to appear. You may need to check your Spam/Junk folder if the email does not appear in your inbox. A supervisor/coach is responsible for ensuring the student receives an email confirmation of ethical approval/waiver. This means a supervisor/coach may need to forward the email notification they have received from Qualtrics directly to their student(s).

Typically, if an email notification from Qualtrics is not received, it is due to an incorrect email address being provided when completing the form. You may need to resubmit the form if there is an error. If after resubmitting the form and correcting email addresses a notification email from Qualtrics does not appear, please contact the Projects Team.

What if I have a question about this form?

- If you have a question about the project requirements of the student's course, contact the Course Leader
- If you have a question about research ethics, contact WMG-FT-SPA@warwick.ac.uk

Q1.2. What programme is your project student(s) enrolled on?
- Full-time Postgraduate Taught programme (e.g. FT MSc)

Q1.3. Are you delegating an ethics outcome to a group student research project, such as an Industry Impact Project?
- No

Q1.4. Is this your first year supervising a taught student project at WMG?
- Yes

Section 1. Student Information for Individual Research Project

All questions must be answered with the correct information to ensure the student receives an email notification of ethical approval/waiver.

Q2.2. What is the student's name? ***Please check this is filled in correctly by copying the information from the student's Tabula profile.***
Guilherme Gomes de Sousa Noronha Jardim

Q2.3. What is the student's Warwick ID number? ***Please check this is filled in correctly by copying the information from the student's Tabula profile.***
5632744

Q2.4. What is the student's Warwick email address? ***Please check this is filled in correctly by copying the information from the student's Tabula profile.***
guilherme.noronha-jardim@warwick.ac.uk

Q2.5. What is the student's project title? ***This information can be copied from the student's ethics form.***
Perceptual Sensitivity to LOD-Induced Animation Artefacts in Self versus Non-Self Avatars

Q2.6. What full-time master's course is the student enrolled on? ***If the course is not listed, return to the first page of this form to select the correct programme.***

- MSc Games Engineering

Section 2. Supervisor/Coach Information

All questions must be answered with the correct information to ensure you receive an email notification of ethical approval/waiver. You can locate information about your staff ID and Warwick email address on your Tabula profile.

Q4.2. What is your name?

David Petrescu

Q4.3. What is your university ID number? ***This can be found on your Tabula profile.***

2470392

Q4.4. What is your Warwick email address? ***Please check this is filled in correctly so you can receive an email notification following submission of this form.***

david-gabriel.petrescu@warwick.ac.uk

As you have selected that this is **your first year supervising a taught student project at WMG**, please be aware that ALL supervisors/coaches must complete the following mandatory training in order to delegate ethical approval/waiver to taught student research:

- Epigeum
- Information Security Essentials
- WMG Supervisor Training on Research Ethics

To support you in your role as a supervisor/coach, please be aware of the following resources:

- WMG guidance for staff on taught student projects
    - Guidance on Ethical Approval of Student Projects
    - Policy for Ethical Approval for Taught Student Projects

- University of Warwick guidance on research projects
    - Research Compliance
    - Research Integrity

Section 3. Confirmations of Research Ethics and Upload of Student Ethics Form

All statements must be responded to in order to complete this form.

Q5.2. I have checked the student ethics form and can confirm that:

- 1. I am satisfied that the student has completed ALL mandatory training as required by the University: **WMG Ethics Training, Epigeum, and Information Security Smart**
- 2. I have discussed the importance of ethical approval with the student(s) and am satisfied that they understand the importance of maintaining high ethical standards when conducting research.
- 3 - FT MSc. I have discussed with the student(s) that data collection must take place whilst they are resident in the UK.
- 4. I have discussed with the student(s) that one of the primary conditions of being granted ethical approval/waiver is the requirement to work with their supervisor/coach to develop their research methods, approach and tools.

Q5.3. Please indicate if any of the following cases has been mentioned in regard to the student's ethics form:

- None of the above

Q6.1. Please upload a PDF record of the student ethics form that you are granting/waiving ethical approval for. ***Note if the student has created multiple versions of the ethics form, it is imperative that only the approved version is uploaded here.***

[Click here]

Q6.2. Please enter the version number of the student ethics form that you are approving. ***This is to ensure the correct ethics form is being approved in case the student has submitted multiple ethics forms.***

1

Q6.3. By submitting this form, I confirm:

- Ethical approval has been **granted** for this research, which uses primary data and/or secondary data that is not publicly available.

Embedded Data:

N/A

Response Summary:

.2 Consent Form

Supervisor Delegated Approval (SDA) Form for WMG Taught Student Projects

This form is NOT for supervisors of cyber security students

Important information:

- This form is for taught student projects ONLY. EngD/PhD/Postgraduate Research students must receive ethical approval from BSREC
- Approval can be delegated to taught student projects that are LOW RISK (click here for information)
- On submission of this form, a notification email will be sent by Qualtrics. The email confirmation of approval/waiver must be included in the student's project submission

What is this form about?
All supervisors/coaches must grant or waive ethical approval for their student's research project BEFORE the student carries out data collection. There are four possible outcomes a supervisor can delegate to a student's ethics form: approve, waive, defer, and not approve.

This form is a legally binding agreement, therefore, when you sign and submit the form you are acknowledging your role as a supervisor in advising the student to maintain high professional standards of ethical conduct as set down in UK legislation. This includes maintaining the ethical principles and guidance provided by WMG and the University of Warwick.

When is the deadline to submit this form?
Supervisors/coaches must submit this form as soon as possible to enable the student to complete their research project and dissertation with sufficient time prior to the deadline. If it is not possible to grant or waive ethical approval, then it is important the supervisor/coach:

- Provides feedback to the student so necessary changes can be made to the ethics form to delegate approval/waiver
- Contacts the WMG Projects Team to inform them of any delays to the student's project completion
- Completes this form selecting the option "Ethical approval is not granted" and provides an explanation for this selection **OR** "Ethical approval requires deferral" and provides an explanation for this selection

Who needs to fill in this form?

The supervisor/coach is responsible for completing this form for each of their project students. A student will NOT be able to complete this form as it is for *supervisor* delegated approval of student research.

What information will be asked?

The form seeks information about the supervisor/coach, the project student, the student ethics form, and confirmations related to ethical approval.

To complete this form, please have the following information ready:

- Student's ID number and email address
- A PDF of the student's ethics form

To generate a PDF of the student's ethics form, print the email from Qualtrics containing the responses to the student ethics form and 'save as PDF'.

How many times can a supervisor submit this form?
This form can be submitted as many times as needed. Students must include the email confirmation of ethical

The personal information we collect and use to conduct this research will be processed in accordance with UK data protection law as explained in the Participant Information Sheet and the Privacy Notice for Research Participants.

By signing this consent form, I declare that I have provided accurate identity details (full name and electronic signature).

_________________________	_________________________	______________
Name of Participant	Signature	Date
_________________________	_________________________	______________
Name of the person taking consent	Signature	Date

1 copy for the participant and 1 copy for the research team (original)

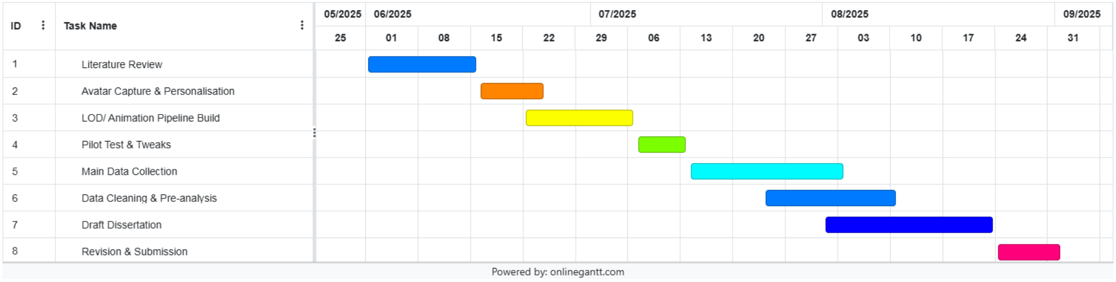


Figure 2: MSc Project Timeline (Gantt Chart)

Bibliography

Abrams, R. A. and Christ, S. E. (2003). Motion Onset Captures Attention. *Psychological Science*, 14(5):427–432.

Bailey, R., Wise, K., and Bolls, P. (2009). How Avatar Customizability Affects Children's Arousal and Subjective Presence During Junk Food-Sponsored Online Video Games. *CyberPsychology & Behavior*, 12(3):277–283. Publisher: Mary Ann Liebert, Inc., publishers.

Benda, B. and Ragan, E. D. (2021). The Effects of Virtual Avatar Visibility on Pointing Interpretation by Observers in 3D Environments. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 50–59. ISSN: 1554-7868.

Carruthers, G. (2008). Types of body representation and the sense of embodiment. *Consciousness and Cognition*, 17(4):1302–1316.

Cummings, J. J. and Bailenson, J. N. (2016). How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology*, 19(2):272–309.

Ennis, C., McDonnell, R., and O'Sullivan, C. (2010). Seeing is believing: body motion dominates in multisensory conversations. *ACM Trans. Graph.*, 29(4):91:1–91:9.

Games, E. (2025a). *MetaHuman Creator and MetaHuman Framework Documentation*. Epic Games.

Games, E. (2025b). *Unreal Engine 5.6 Documentation*. Epic Games. Version 5.6.

Gescheider, G. A. (2013). *Psychophysics: The Fundamentals*. Psychology Press, 3 edition.

Hautus, M. J., Macmillan, N. A., and Creelman, C. D. (2021). *Detection Theory: A User's Guide*. Routledge, 3 edition.

Hodgins, J., Jörg, S., O'Sullivan, C., Park, S. I., and Mahler, M. (2010). The saliency of anomalies in animated human characters. *ACM Trans. Appl. Percept.*, 7(4):22:1–22:14.

Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203.

Jardim, G. N. (2025). lod-detection-thresholds: Analysis code and data for lod detection thresholds.

Kang, S., Yoon, B., Kim, B., and Woo, W. (2022). Effects of Avatar Face Level of Detail Control on Social Presence in Augmented Reality Remote Collaboration. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pages 763–767. ISSN: 2771-1110.

Kenny, S., https://orcid.org/0000-0002-7908-1506, View Profile, Mahmood, N., View Profile, Honda, C., View Profile, Black, M. J., View Profile, Troje, N. F., and View Profile (2019). Perceptual Effects of Inconsistency in Human Animations. *ACM Transactions on Applied Perception*, 16(1):1–18. Publisher: Association for Computing Machinery.

Kenny, S., Mahmood, N., Honda, C., Black, M. J., and Troje, N. F. (2017). Effects of animation retargeting on perceived action outcomes. In *Proceedings of the ACM Symposium on Applied Perception*, pages 1–7, Cottbus Germany. ACM.

Kingdom, F. A. A. and Prins, N. (2016). *Psychophysics: A Practical Introduction*. Elsevier Science & Technology.

Knodt, J. (2025). Single Edge Collapse Quad-Dominant Mesh Reduction. *ACM Transactions on Graphics*, 44(4):1–23.

Lee, Y. J., , and Ji, Y. G. (2025). Effects of Visual Realism on Avatar Perception in Immersive and Non-Immersive Virtual Environments. *International Journal of Human-Computer Interaction*, 41(7):4362–4375. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/10447318.2024.2351713.

Lombardi, S., Saragih, J., Simon, T., and Sheikh, Y. (2018). Deep appearance models for face rendering. *ACM Transactions on Graphics*, 37(4):1–13.

Luebke, D., Reddy, M., Cohen, J. D., Varshney, A., Watson, B., and Huebner, R. (2002). *Level of Detail for 3D Graphics*. Elsevier. Google-Books-ID: uiz6IKAVxP8C.

Lugrin, J.-L., Latt, J., and Latoschik, M. E. (2015). Avatar anthropomorphism and illusion of body ownership in VR. In *2015 IEEE Virtual Reality (VR)*, pages 229–230. ISSN: 2375-5334.

Maselli, A. and Slater, M. (2013). The building blocks of the full body ownership illusion. *Frontiers in Human Neuroscience*. Place: Lausanne, Switzerland Publisher: Frontiers Research Foundation Section: Original Research ARTICLE.

Mori, M., MacDorman, K. F., and Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100.

Nguyen, D. V., Hien, T. T., and Huong, T. T. (2024). A Subjective Quality Evaluation of 3D Mesh With Dynamic Level of Detail in Virtual Reality. In *2024 IEEE International Conference on Image Processing (ICIP)*, pages 1225–1231, Abu Dhabi, United Arab Emirates. IEEE.

Ogawa, N., Narumi, T., and Hirose, M. (2018). Object Size Perception in Immersive Virtual Reality: Avatar Realism Affects the Way We Perceive. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 647–648.

Petrescu, D., Warren, P. A., Montazeri, Z., and Pettifer, S. (2023). Velocity-Based LOD Reduction in Virtual Reality: A Psychometric Approach. arXiv:2301.09394 [cs].

Rosenholtz, R. (2016). Capabilities and Limitations of Peripheral Vision. *Annual Review of Vision Science*, 2(1):437–457.

Runeson, S. and Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person-and-action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, 112(4):585–615. Place: US Publisher: American Psychological Association.

Slater, M., Spanlang, B., Sanchez-Vives, M. V., and Blanke, O. (2010). First Person Experience of Body Transfer in Virtual Reality. *PLoS ONE*, 5(5):e10564.

Strasburger, H., Rentschler, I., and Juttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5):13.

Sun, X., Xu, Y., Dingliana, J., and O'Sullivan, C. (2025). Evaluating CrowdSplat: Perceived Level of Detail for Gaussian Crowds. arXiv:2501.17085 [cs].

Waltemate, T., Gall, D., Roth, D., Botsch, M., and Latoschik, M. E. (2018). The Impact of Avatar Personalization and Immersion on Virtual Body Ownership, Presence, and Emotional Response. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1643–1652.

Wichmann, F. A. and Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. 63(8):1293–1313. Company: Springer Distributor: Springer Institution: Springer Label: Springer Publisher: Springer-Verlag.

Yamac, G., O'Sullivan, C., and Neff, M. (2024). Understanding the Impact of Visual and Kinematic Information on the Perception of Physicality Errors. *ACM Trans. Appl. Percept.*, 22(1):1:1–1:30.

Yantis, S. and Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 10(5):601–621.