# VR OBSERVATION FEAT



Ayanokoji is inside a full 360-degree VR Tokyo with moving crowds, layered audio, depth, and motion, while suffering motion sickness that disrupts vestibular balance. Despite that, he must detect unnatural points, meaning errors subtle enough that they are not obvious visual glitches but violations of realism, behavior, physics, or timing. Because pain is the correction method,

failure is not allowed. This forces nonstop observation without rest, blinking minimization, and sustained precision.

All the children stood up and put on large goggles at the same time.

Our vision went black, but soon the screen lit up and the program was displayed, and it began after a few moments.

"The curriculum will now focus on Japan, whereas in the past we've studied American cities such as New York and Hawaii. First, we'll start with public transportation."

This was the basic premise of the course. It introduced a world that wasn't just a White Room.

This was still learning time, and children were told early on that they won't leave this place until they become adults.

The virtual console reproduced the same outside scenery in 360 degrees with such quality that it could be mistaken for the real thing, and the sound was combined with the visuals to create a sense of presence. Even the people passing by were reproduced, showing a businessman in a suit, an old man with a cane, an elderly woman trying to get into a cab, and other street scenes.

Of course, children were also present, but unlike the reality outside, they didn't appear to be playing or having fun at all; instead, they showed inorganic, machine-like movements.

We learned the history and structure of the world so that one day, when we go out into the outside world, we'll be able to adapt to it without problems.

I knew it was necessary, but I had a problem with this way of learning.

One of the reasons why I disliked it was because it was accompanied by an indescribable feeling of discomfort.

It's what was commonly described as 3D motion sickness.

It's possible that the brain misperceives it as a hallucination if the balance between visual perception and the semicircular canals are incorrect.

There's no way to stop the sickness by individual power alone, and the only way would be to let the brain learn over time.

It wasn't so hard that it was impossible to continue, but it was the reason why I didn't like it.

Of course, the virtual console wasn't only used as a device to perceive the outside world visually but also as a tool to train observation and insight.

We were asked to detect unnatural points in the views that unfolded in various locations.

If what we pointed out was wrong or the unnatural point itself couldn't be found, the instructors gave us unrelenting guidance.

The methods of guidance varied, but it mainly consisted of those that caused pain to the students themselves.

That's why we used our eyes to thoroughly observe, not even sparing the blink of an eye.

The more we feared for our lives, the more our senses sharpened and we began to see things that we couldn't see before.

"Next, let's take a walk in Tokyo on the virtual console."

As we virtually walked through Tokyo, the screen suddenly went dark.

The instructors' voices that I was listening to stopped, and I was engulfed in silence.

"Everyone take off your goggles."

The voice came from inside the room, not through the microphone, and we all followed the instruction at once.

"There's an equipment issue. That's it for today's virtual console lesson. We still have less than half an hour before the next curriculum, so please stay here."

With those instructions, the goggles in everyone's hands were retrieved.

"Stand by..."

# What he is doing cognitively

He is running three processes at once.
Global scanning. His eyes sweep a 360-degree environment while maintaining a mental map of everything already seen. Humans normally hold 3 to 5 tracked objects. He is holding dozens.
Local detail inspection. He zooms into individuals and objects to check micro-features like gait rhythm, posture shifts, reaction delay, and movement smoothness.

# Calculations

Let’s quantify what is on screen.

**A realistic Tokyo street scene would easily include:**

30 to 60 visible people
10 to 20 moving vehicles
50 to 100 static environmental objects
Continuous audio sources

We will low ball hard.

Visual sampling rate

Human eyes make about 3 to 5 saccades per second. With training and fear, assume 5 saccades per second.

Each saccade samples a new region.

That gives:

5 focused checks per second

Over 10 minutes:
5 × 600 = 3,000 focused visual inspections

But that is only focal vision.

Peripheral vision is always active.

**LOW BALL (no cognitive limit, conservative task load)**

Assume he actively tracks:
20 entities at once
Each entity requires 5 variables:
Position
Velocity
Direction
Behavior pattern
Context consistency

That is: 20 × 5 = 100 variables
If each variable updates at just 2 times per second: 100 × 2 = 200 updates per second

This means his brain is:
Updating 200 live variables per second
Comparing each against a realism model
While sick and under threat. This alone is already far beyond normal human capacity.

Low ball result: He is performing 200 real-time validation checks per second with no cognitive fatigue.

**MID BALL (expanded realism checking)**
**Now include prediction.**

Each entity is not just observed but projected forward.
Assume:
30 entities
6 variables each
Prediction window of 1 second ahead
Updated 3 times per second

That is: 30 × 6 × 3 = 540 predictive checks per second

Each check includes:
What should happen
What actually happens

Whether deviation is acceptable or unnatural
This means he is running a half-thousand comparisons per second, continuously.
And remember, this is a child.

**HIGH BALL (no cognitive limits, full exploitation)**

Now we go all in.
He is not tracking entities. He is tracking systems.
Assume:
50 people
15 vehicles
40 environmental anchors Total: 105 elements

Each element has:
Spatial data
Temporal data
Behavioral logic
Interaction rules

Let's say 8 variables per element.
105 × 8 = 840 variables

If each variable updates at 5 times per second: 840 × 5 = 4,200 variable updates per second

Now add cross-comparisons.

If just 10 percent of these variables must be cross-checked with others: 4,200 × 0.1 = 420 relational checks per second

Total operations per second: 4,200 + 420 = 4,620 checks per second

The smallest details he must notice:

1. Step timing differences under 100 milliseconds
2. Sound lag mismatches under 50 milliseconds
3. Gait symmetry errors under 2 percent
4. Behavioral stiffness across repeated cycles
5. Crowd flow gradients that feel "off" but look fine

NPC reactions that are correct individually but wrong collectively

Missing any of these results in punishment.

# Final interpretation



**Low ball, he is already operating at hundreds of real-time checks per second.**

**Mid ball, he is predicting human and environmental behavior faster than it unfolds.**

**High ball, he is effectively running a city-scale simulation validator in his head, nonstop. He is not observing Tokyo. He is stress-testing reality.**

And the most important part? This was not a special feat. This was class.

**NOW WE HAVE TO GO TO PRIME KOJI AT 14 YRS OLD**

**First, what changes from age 5 to 14?**

Three things fundamentally scale.

1. **Sensory resolution improves**

Visual acuity, contrast sensitivity, motion discrimination, and micro-timing detection all mature. He literally sees finer differences and faster changes.

2. **Internal modeling depth increases**

At 5, he detects “unnatural.”
At 14, he detects why it is unnatural, how it propagates, and what it implies.

3. **Parallelization becomes extreme**

What was once effortful becomes background processing. Conscious attention is freed up for higher-order inference.

So the same task explodes in magnitude.

**Baseline assumption for Prime (age 14)**
**By this point:**

VR environments are more complex

Instructors deliberately hide flaws

Punishment threshold is higher

Sessions are longer

Detection speed expectations are stricter
Meaning the task difficulty increases faster than his age.

**LOW BALL PRIME (still conservative)**

We take the child low ball and scale it only by:

Improved perception

Better automation

Child low ball: 200 variable checks per second

Prime conservative multiplier: ×5

Why only 5? Because we are being insulting on purpose.

200 × 5 = 1,000 checks per second

**This already means:**

He can track roughly 100 entities

Each with 10 live variables

Updating in real time

Low ball prime Ayanokoji is already operating at four figures per second, continuously, without fatigue.

**MID BALL PRIME (realistic growth)**

Now we account for nine years of nonstop exposure.

**At this point:**

Pattern libraries exist

Prediction replaces observation

Deviations are caught before they visually manifest

Assume:

150 entities

8 variables per entity

Update rate: 4 times per second
150 × 8 × 4 = 4,800 updates per second

Now add relational inference.

Even if only 25 percent require.

cross-validation: 4,800 × 0.25 = 1,200 relational checks per second.

Total: 4,800 + 1,200 = 6,000 operations per second

Mid ball prime Ayanokoji is doing six thousand real-time environmental validations per second.

That is already beyond any human benchmark.

**HIGH BALL PRIME (full exploitation, no restraint)**

Now we assume White Room instructors are actively trying to beat him.

**That means:**

No obvious errors

Multi-layered inconsistencies.

Errors are only detectable through system-level logic.

He is no longer tracking objects.

He is tracking probability fields.

**Assume:**

250 dynamic elements

10 variables each

Updated at 5 times per second

250 × 10 × 5 = 12,500 variable updates per second

Now add higher-order modeling.

**If even 40 percent require:**

Temporal consistency checks

Behavioral intent validation

Crowd dynamics coherence

12,500 × 0.4 = 5,000 higher-order checks per second.

Total: 12,500 + 5,000 = 17,500 cognitive operations per second

# NOW FOR PROCESSING SPEED (😭✌️)

**Step 1. Define the visual load (absolute low ball)**

We take the VR Tokyo street scene.

**Conservative assumptions**

Field of view processed at a time: 120° horizontal, not full 360

Effective visual resolution processed consciously: 60 pixels per degree

This is lower than human foveal vision
Vertical FOV: 60°.

**So visual input per frame:**

120 × 60 = 7,200 pixels wide
60 × 60 = 3,600 pixels tall

Total per frame: ~26 million pixels
That is already low, but we go even lower.

**Step 2. Frame rate (low ball)**

VR minimum to avoid motion sickness is about 60 fps

We will assume he only meaningfully processes 30 fps

So per second: 26,000,000 × 30 = 780,000,000 pixels per second
Under 1 billion. Floor level.

**Step 3. What counts as “processing” per pixel?**

We will not assume full semantic interpretation. Just binary anomaly detection.

**Each pixel must be:**

Compared to surrounding pixels

Checked for motion consistency

Checked for texture or lighting mismatch

We low ball this to 3 comparisons per pixel

So operations per second: 780,000,000 × 3 ≈

2.34 billion micro-operations per second
This is just for raw perception.

**Step 4. Object-level processing**
**Now objects.**

Again, low ball.

People visible at any moment: 15

Vehicles: 5

Static objects (signs, poles, doors): 30

Total objects: 50

**Each object requires:**

Shape consistency

Motion logic

Behavioral normality check

We low ball to 10 variables per object

So per frame: 50 × 10 = 500 checks

At 30 fps: 15,000 object-level checks per second.

Negligible compared to pixel load, but still real.

**Step 5. Time to detect an anomaly**
**Here is the killer part.**

The instructors punish misses, meaning detection must happen before scene changes.
Scene coherence window in VR walking speed: ~200 milliseconds

That means all of this processing must happen in:

0.2 seconds

So we divide.

**Pixel operations in 0.2s: 2.34 billion ÷ 5 around and equal to 468 million operations in 200 ms**

That is:

~2.34 billion ops per second sustained ~468 million ops per reaction window
And this is low balling everything.

At age fourteen, this observation feat becomes fundamentally different from the same task at age five, not because the environment changes, but because the way his brain handles information has been radically compressed through nine uninterrupted years of White Room conditioning. The key shift is that he no longer perceives the scene as a collection of separate details. Instead, he recognizes structured patterns instantly. Human movement, spacing, posture, timing, and environmental consistency are no longer processed one by one. They are processed as unified packets. This drastically reduces cognitive overhead while increasing

accuracy. On top of that, his attention is no longer evenly distributed. He anticipates where irregularities are statistically most likely to appear, meaning most of his processing power is focused on high value zones rather than wasted on neutral background information.

By fourteen, his reaction window also tightens. The fear conditioning, punishment based correction, and repetition push perceptual decision making well below average adult reaction times. He is not reacting after the fact. He is recognizing and validating anomalies almost as soon as the visual input stabilizes. This allows him to scan dense urban scenes in fractions of a second without panic, hesitation, or loss of detail. The result is not superhuman vision, but superhuman interpretation speed. His eyes remain human. His brain is simply no longer bottlenecked by hesitation, indecision, or sequential processing.

What makes this feat extreme is that it is sustainable. This is not a burst of focus or a lucky moment. It is a trained baseline state. He can repeat this level of observation continuously under stress, motion sickness, and threat of punishment. That consistency is what elevates the feat from impressive perception to elite level cognitive processing. By fourteen, the same task that once required conscious effort has become automatic, faster, and far denser in informational output.

# Calculations

Baseline at age 5 (mid reference):

230 object-features processed in 0.25 seconds = 920 object-features per second

Age 14 scaling assumptions:

Improved chunking efficiency

Reduced wasted attention

Shorter perceptual decision window

**Low ball:**

230 objects in 0.15 seconds

230 ÷ 0.15 ≈ 1,530 objects per second

**Mid ball:**

600 objects in 0.10 seconds

600 ÷ 0.10 = 6,000 objects per second

**High ball:**

1,200 objects in 0.08 seconds

1,200 ÷ 0.08 = 15,000 objects per second

# Clean Final Scaling Summary

**Age: 14**

Objects processed per scene:

Low ball ≈ 230

Mid ball ≈ 600

High ball ≈ 1,200

**Time required:**

Low ball ≈ 0.15 seconds

Mid ball ≈ 0.10 seconds

High ball ≈ 0.08 seconds

**Processing rate:**

Low ball ≈ 1.5k objects per second

Mid ball ≈ 6k objects per second

High ball ≈ 15k objects per second

This remains within physical sensory limits while demonstrating extreme cognitive compression, parallel processing, and anticipatory filtering.